KØBENHAVNS UNIVERSITET
INSTITUT FOR NATURFAGENES DIDAKTIK

Maximum Likelihood Estimation

Subject Matter Didactic Analysis and Didactical Design

Cæcilie Bøje Pedersen

Speciale i Matematik & Didaktik

Vejledere:
Carl Winsløw

May 2025

IND's studenterserie nr. 129, 2025

INSTITUT FOR NATURFAGENES DIDAKTIK, www.ind.ku.dk
Alle publikationer fra IND er tilgængelige via hjemmesiden.

IND's studenterserie

90. Maria Anagnostou: Trigonometry in upper secondary school context: identities and functions (2020)
91. Henry James Evans: How Do Different Framings Of Climate Change Affect Pro-environmental Behaviour? (2020)
92. Mette Jensen: Study and Research Paths in Discrete Mathematics (2020)
93. Jesper Hansen: Effekten og brugen af narrative læringsspil og simuleringer i gymnasiet (2020)
94. Mie Haumann Petersen: Bilingual student performance in the context of probability and statistics teaching in Danish High schools (2020)
95. Caroline Woergaard Gram: "Super Yeast" - The motivational potential of an inquiry-based experimental exercise (2021)
96. Lone Brun Jakobsen: Kan man hjælpe elevers forståelse af naturvidenskab ved at lade dem formulere sig om et naturvidenskabeligt emne i et andet fag? (2021)
97. Maibritt Oksen og Morten Kjøller Hegelund: Styrkelse af motivation gennem Webinar og Green Screen (2021)
98. Søren Bystrup Jacobsen: Peer feedback: Fra modstand til mestring? (2021)
99. Bente Guldbrandsen: Er der nogen, som har spurgt en fysiklærer? (2021)
100. Iben Vernegren Christensen: Bingoplader i kemiundervisningen – en metode til styrkelse af den faglige samtale? (2021)
101. Claus Axel Frimann Kristinson Bang: Probability, Combinatorics, and Lesson Study in Danish High School (2021)
102. Derya Diana Cosan: A Diagnostic Test for Danish Middle School Arithmetics (2021)
103. Kasper Rytter Falster Dethlefsen: Formativt potentiale og udbytte i Structured Assessment Dialogue (2021)
104. Nicole Jonasen: A diagnostic study on functions (2021)
105. Trine Nørgaard Christensen: Organisatorisk læring på teknisk eux (2021)
106. Simon Funch: Åben Skole som indgang til tværfagligt samarbejde (2022)
107. Hans-Christian Borggreen Keller: Stem som interdisciplinær undervisningsform (2022)
108. Marie-Louise Krarup, Jakob Holm Jakobsen, Michelle Kyk & Malene Hermann Jensen: Implementering af STEM i grundskolen (2022)
109. Anja Rousing Lauridsen & Jonas Traczyk Jensen: Grundskoleelevers oplevelse af SSI-undervisning i en STEM-kontekst. (2022)
110. Aurora Olden Aglen: Danish upper secondary students' apprehensions of the equal sign (2023)
111. Metine Rahbek Tarp & Nicolaj Pape Frantzen: Machine Learning i gymnasiet (2023)
112. Jonas Uglebjerg: Independence in Secondary Probability and Statistics: Content Analysis and Task Design (2023)
113. Hans Lindebjerg Legard: Stopmotion som redskab for konceptuel læring. (2023)
114. Caroline Woergaard Gram & Dan Johan Kristensen: The ice algae Ancylonema as icebreakers: A case study on how the international Deep Purple Research Project can create meaningful outreach in Greenland. (2023)
115. Julie Sloth Bjerrum: 'KLIMA HISTORIER' The Art Of Imagining A Green Future. (2023)
116. Emilie Skaarup Bruhn: Muligheder og udfordringer ved STEM-undervisning (2024)
117. Milla Mandrup Fogt: Undersøgelsesbaseret undervisning i Pascals trekant (2024)
118. Mille Bødstrup: P-hacking (2024)
119. Nynne Milthers & Amanda Wedderkopp: Inquiry of the Past and Reflection on the Present: Teaching Rigour and Reasoning in Area Determination through Authentic Historical Sources (2024)
120. Pelle Bøgild: Med kroppen ind i fysikken (2024)
121. Anne Jensen & Charlotte Puge: Modellering som bro mellem teoretisk viden og praktisk laboratoriearbejde (2024)
122. Anne Sofie Berendt: ChatGPT som formativ feedbackgiver på gymnasieelevers design af biologieksperimenter (2024)
123. Christian Gothard Rix, Jacob Broe & Mads Kiilerich Kirdan: Anvendelse af ChatGPT3.5 i gymnasiet (2024)
124. Astrid Junker Andersen & Ditte Zimmermann Bramow: Faglig læsning af multimodale tekster i biologi (2024)
125. Mette Elsnab Olesen: En kropslig tilgang til STEM - Et casestudie omhandlende grundskolelæreres opfattelser og anvendelser af Embodied Cognition-perspektiver i STEM-undervisning (2024).
126. Victor Lynge Nielsen: Probabilistic modeling and AI in upper secondary mathematics (2025)
127. Mads Fencker: Epistemic Value Pluralism in the Practice of Stochastic Analysis (2025)
128. Josefine Dixen Zwisler: The Mathematics of Map Projections (2025)
129. **Cæcilie Bøje Pedersen: Maximum Likelihood Estimation (2025)**

__IND's studenterserie__ omfatter kandidatspecialer, bachelorprojekter og masterafhandlinger skrevet ved eller i tilknytning til Institut for Naturfagenes Didaktik. Disse drejer sig ofte om uddannelsesfaglige problemstillinger, der har interesse også uden for universitetets mure. De publiceres derfor i elektronisk form, naturligvis under forudsætning af samtykke fra forfatterne. Det er tale om studenterarbejder, og ikke endelige forskningspublikationer.
Se hele serien på: www.ind.ku.dk/publikationer/studenterserien/

Abstract

This is a thesis within the didactics of statistics and mathematics that investigates how to introduce likelihood functions and maximum likelihood estimation in Danish upper secondary school. The thesis examines how the topic can be meaningfully communicated to strengthen the content of statistics and create links between statistics, probability and mathematics. Since likelihood functions and maximum likelihood estimation as topic is largely unknown in the context of upper secondary education, and that very little didactic literature exists on the topic, the purpose is to carry out a new didactic transposition. This is done by analysing the content of likelihood functions and maximum likelihood estimation and by developing a textbook chapter. The analysis consists of a subject matter didactic analysis conducted within the theoretical framework known as *'Stoffdidaktik'*. The analysis use the tools *'aspects'* and *'Grundvorstellungen'* to describe the content and explore which competencies students might acquire to understand the topic. Furthermore, the analysis identify subject-didactic perspectives used to legitimize likelihood functions and maximum likelihood estimation as knowledge to be taught. In total, four subject-didactic perspectives are identified that support the teachability of the topic. Secondly, to organize the didactical design, an analysis of the textbook chapter using the theoretical framework of the *'Anthropological Theory of the Didactic'* is conducted. This analysis consists of an institutional analysis of the conditions that must be considered when implementing likelihood functions and maximum likelihood estimation in teaching practice. It was found that the topic places high demands on both teachers and students in terms of qualifications and prior knowledge, due to its theoretical nature. Finally, the analysis includes an a priori analysis on the praxeological organization and the task design. This analysis offers a proposal for how a didactical design on likelihood functions and maximum likelihood estimation could be introduced in upper secondary school. However, it also identifies several heuristic difficulties related to task design, which teachers should be aware of when using the textbook chapter.

Table of Contents

1 Introduction

The background for this thesis is my professional interest for statistics and my experience as an upper secondary school teacher during the last three years. My wish has been to write a thesis that combines my passion for teaching and lesson planning, and my master in statistics. Over the past four months, this has taken shape as a theoretical didactic thesis about the method of Maximum Likelihood Estimation (MLE).

I have chosen my topic MLE because of my education in theoretical statistics and to explore how an exciting statistical topic can be incorporated into an upper secondary school context and communicated to students. **Maximum Likelihood Estimation** is a method that uses a sample to estimate the parameters in an assumed statistical model. Estimation is one of the most fundamental concepts in statistics, where it is used to calculate or guess a value. The MLE method was invented by the British statistician and geneticist Sir Ronald Aylmer Fisher (1890-1962) at the beginning of the 20th century and today the method is widely used in machine learning and applied statistics [Efron and Hastie, 2016].

It is my belief that MLE is not a typical teaching topic in upper secondary school. It has actually been quite a journey to investigate whether any research in the didactics of mathematics and statistics has been done on MLE – and whether any teaching materials about MLE exist for upper secondary school. After thorough investigation, very few didactical papers aimed at upper secondary school on MLE have been found. For example, a search in the International Association for Statistical Education (IASE) website yields only nine hits from their collected conferences and journals. All the nine hits are aimed at the university level [IASE, 2025]. Therefore, an essential part of my thesis has been to figure out how to organize and design a teaching material on MLE for upper secondary school, since there is very limited inspiration and didactical literature to be found. Furthermore, it is important to argue why it makes sense to teach MLE in upper secondary school at all. My goal of this thesis is to contribute to a new didactic transposition within statistics by developing a didactical design on likelihood functions and MLE aimed at upper secondary school.

In Danish upper secondary school, statistics is not a subject of its own, but it appear in school mathematics as a subfield [Børne- og Undervisningsministeriet, 2024a]. Likewise statistics is often described as applied mathematics, as it relies heavily on the use of probability models to describe stochastic phenomena. Despite the important role of probability, the key element in statistics is data, and statistics involves all the processes related to data: collecting data, constructing models, data analysis and finally the interpretation [Ekstrøm et al., 2017].

Statistics appear in one of the three branches in national mathematics syllabus: (1) trigonometry and vectors, (2) functions and calculus, and (3) probability and statistics [Børne- og Undervisningsministeriet, 2024b]. The field of statistics as a branch creates the institutional framework for statistical education in Danish upper secondary school: what statistical content should be taught and what are the possibilities for an optional topic in statistics? In my experience as teacher, you are busy getting through all the core content, leaving little time for optional or supplementary content. In 2024, mathematics got a new syllabus stating that 10 percent of the total teaching must be spent on supplementary content [Børne- og Undervisningsministeriet, 2024a]. This revision makes it possible to allocate time for optional topics, such as my proposal of MLE to be taught in upper secondary school. The aim of this thesis is not to discuss the role of statistics in the syllabus, or why the 10 precent could advantageously be allocated to topics in statistics, but to investigate how a concrete teaching material could be organized and designed introducing the topic of MLE to students. My target group of the design is A-level mathematics students.

The main research question of this thesis is: *How can likelihood functions and maximum likelihood estimation be introduced in Danish upper secondary school mathematics in a way that strengthens the content in statistics and establishes connections between concepts in statistics, probability and mathematics?* In section 3, I will further detail the research question into a number of sub-questions.

The foundation is that such introduction of MLE to upper secondary school students is possible due to the new syllabus from 2024. To answer the research questions, I have developed a textbook chapter on MLE, which could potentially be used as the supplementary topic of the syllabus. To develop the chapter, I have conducted extensive preparatory work.

First, I introduce the theoretical framework (section 2), the history of likelihood (section 4) and I give a brief introduction to probability (section 5). Secondly, I use different types of didactical analysis to approach the didactic transposition of likelihood functions and MLE.

A subject matter didactic analysis on likelihood functions and MLE is conducted (section 6) within the theoretical framework called *'Stoffdidaktik'*. It has been used to organize likelihood functions and MLE by subject-didactic tools, exploring relevant concepts interconnected to MLE. The subject matter didactic analysis has also been used to identify subject-didactic perspectives of likelihood functions and MLE that strengthens the content of statistics and connects concepts in statistics, probability and mathematics.

A didactical design is developed (section 7) consisting of a institutional analysis and a priori analysis, which is part of the research methodology known as didactic engineering. The institutional analysis includes considerations and design choices regarding how the textbook chapter can be implemented in a teaching practice by institutional means. To organize and design the textbook chapter, I have performed an a priori analysis. This analysis includes task design: what is the goal of the tasks, what can students do to solve the tasks and why students choose the strategies as they do. In the didactical design, the Anthropological Theory of the Didactic (ATD) is used as theoretical framework, which is applied to address epistemological and institutional questions in developing my didactical design. For example considering the institutional constraints and conditions relevant for the design or working with the epistemological organization of knowledge in relation to practice and theory. In the a priori analysis the framework is applied to develop an ATD-based reference model for likelihood functions and MLE, which is used to structure the textbook chapter in tasks that students can solve and techniques and technologies students can use in their solution.

Finally, a discussion is conducted (section 8), where I address some of the limitations and contributions of the didactical design, and possible improvements of my investigation.

Despite the rather unusual and theoretical topic (likelihood functions and MLE) in my thesis, I have a great hope that the didactical design will be used in some shape or form and that this thesis can open a conversation about the content of statistics in an upper secondary school context.

2 Theoretical Framework

2.1 Anthropological Theory of the Didactic

As background for analyzing statistical content in Danish upper secondary school mathematics it seems crucial to study the role of statistical education in perspective to the real-life teaching practices and the knowledge related to MLE. To achieve this, a didactic program of *Yves Chevallard (1946–)* is used as a theoretical framework. Chevallard introduced 'the Theory of Didactic Transpositions' in the eighties and later expanded it to 'the Anthropological Theory of the Didactic (ATD)' in the nineties. In the center of his theories is human practices, discourses and institutions related to mathematical knowledge [Winsløw, 2011].

In this view knowledge-related activities such as *"Doing, teaching, learning, diffusing, creating, and transposing mathematics"* [Bosch and Gascón, 2014, p. 68] must be seen in relation to the institutions they are taking place in, and it is the role of didactics to cover this institutional condition [Bosch and Gascón, 2014, p. 68]. In ATD, institutions should be understood widely as structures and systems related to didactic phenomena. One of the key concepts of Chevallard's theories is *didactic transpositions*, which is

> *"..the transformations an object or a body of knowledge undergoes from the moment it is produced, put into use, selected, and designed to be taught until it is actually taught in a given education institution."*
> [Chevallard and Bosch, 2014, p. 214]

In other words, didactic transposition theory is about knowledge circulating between different institutions and actors in a society. The process of didactic transposition involves four types of knowledge; scholarly knowledge, knowledge to be taught, taught knowledge and learnt knowledge. The theory is especially concerned about the evolution and changes to this process, which is sketched in figure 1 [Bosch and Gascón, 2014]. Further, the theory distinguishes between two types of transpositions: *internal* (inside the classroom) and *external* (outside the classroom) [Winsløw, 2011].

Regarding internal didactic transpositions, it is crucial what is being taught, *"taught knowledge"*, what is being adopted from the syllabus, and what is transmitted to the students, as well as what the students actually learn, *"learnt knowledge"*. External transposition concerns the connections between the the microcosm of the classroom and the impact of the outside world. These transpositions are the institutional framework around education containing *the noosphere* and scholarly knowledge [Winsløw, 2011]. Scholarly knowledge is knowledge produced for and by scientists, professors and students at the universities. Thus, scholarly knowledge has a lot of integrity, but its formalization and complexity is rarely suited for pre-academic education. In the didactic process, the noosphere selects and intermediates scholarly knowledge to adapt the level of school education as "knowledge to be taught". This sphere includes textbooks writers, teachers, the Ministry of Children and Education, politics, public opinions and more on [Chevallard and Bosch, 2014].

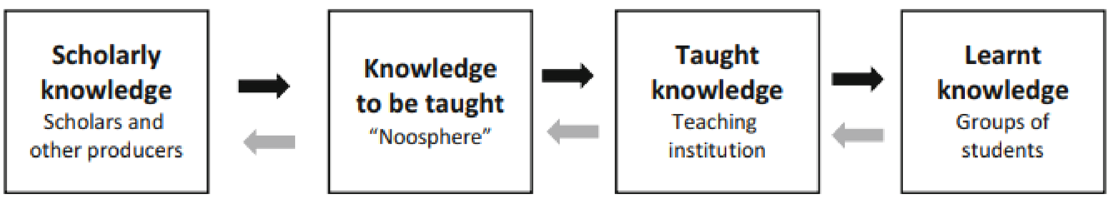


Figure 1: Diagram on the process of didactic transposition [Bosch and Gascón, 2014].

The didactic transposition diagram is a tool for analyzing conditions for didactic phenomena and for developing a teaching practice — its application to this thesis is inspired by Wan Kang and Jeremy Kilpatrick's article *"Didactic Transposition in Mathematics Textbooks"* [Kang and Kilpatrick, 1992]. ATD is used for preparatory work on teaching MLE as a new didactical transposition. The goal is to transform MLE into "knowledge to be taught."

— The first part of the didactical transposition consists of organizing MLE as an integrated whole. According to Kang and Kilpatrick the topic must be derived from scholarly knowledge and given a "coherent theoretical assemblage" [Kang and Kilpatrick, 1992].

— Next, MLE as "knowledge to be taught" must be legitimized as knowledge to be used. I must be able to argue why MLE should be taught and how students can apply it. Hence, subject-didactic arguments must be established to justify the didactic transposition.

— According to Kang og Kilpatrick a textbook chapter is a declared body of knowledge. In the didactic design, MLE must be declared, meaning it must be broken into parts and reconstructed as "knowledge to be taught." In other words, the content must simply be presented in a new way that makes sense to upper secondary school students. This involves identifying structural conditions and constraints that hinder MLE as "knowledge to be taught" and taking these into account when developing textbooks [Kang and Kilpatrick, 1992].

2.2 Praxeologies

Another design tool of ATD is to describe knowledge-related practice and organization in terms of praxeologies. In ATD, a *praxeology* $\mathscr{P}$ is a human practice that can solve mathematical tasks. A praxeology is denoted as a 4-tuple $\mathscr{P} = [T/\tau/\theta^*/\Theta^*]$ [Winsløw, 2011]. This thesis uses a $*$-notation to avoid confusion with the parameter θ in a statistical model. Chevallard divides mathematical activity into two blocks, which identify the practical part of a task and the knowledge-based part of a task. The practical block $[T/\tau]$ consists of types of tasks T and the techniques τ that can be used to solve them. The knowledge block $[\theta^*/\Theta^*]$ consists of technology θ^*, which includes elements that directly apply to solutions such as a definition or a theorem. The second part is theory Θ^*, which provides a deeper and more integrated explanation behind tasks [Barbe Farre et al., 2005]. For example, a theory could be probability theory or number theory. So, a praxeology is simply a theoretical model that identifies and connects practice and knowledge of a task.

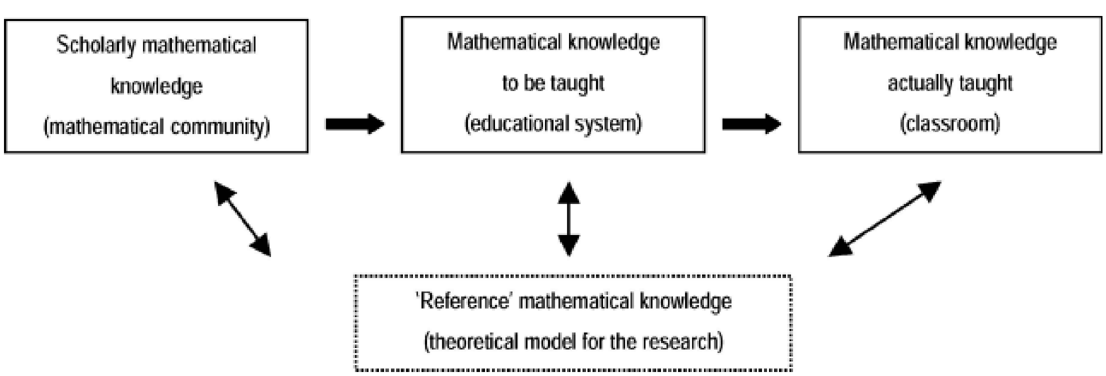


Figure 2: Reference Model of Mathematical Organisation [Barbe Farre et al., 2005]

The tools of didactic transpositions and praxeologies are interconnected, partly because praxeologies cannot be viewed in total isolation from institutional conditions and constraints [Barbe Farre et al., 2005]. In Figure 2, the praxeology-theoretical reference model is shown in relation to the process of didactic transposition. In this thesis, praxeologies is used as design tool for organizing knowledge related to MLE. A praxeologic reference model on MLE is build using the notation $\mathscr{P} = [T/\tau/\theta^*/\Theta^*]$, which the design of the textbook chapter should be based on. My own reference model on MLE developed in this thesis is denoted $\mathscr{P}_{ref}$.

2.3 Subject Matter Didactics

The second theoretical framework used is subject matter didactic, *'Stoffdidaktik'*, which was a school of thought in mathematical education research in the German-speaking countries from the late 1960s to the 1980s [Hofe and Blum, 2016]. This approach in mathematical education research focus on the analysis of mathematical concepts in relation to their content and conceptual understanding [Hußmann et al., 2016]. The approach had a particular purpose of simplifying mathematics education to make it more accessible and understandable for students at the primary and secondary school levels [Hofe and Blum, 2016]. One of the most central figures within the school was the German mathematics educator Arnold Kirsch (1922–2013). He described ways to implement "accessibility" in mathematics education. For example *"making accessible by including the "surroundings" of mathematics"* or *"making accessible by changing the mode of representation"* [Vohns, 2016, p. 214]. Kirsch's notion of "surroundings of mathematics" can be understood as a broader perspective on mathematics, which connects mathematical tasks to real world examples and contextualize it [Vohns, 2016]. In appendix C, I have made a concept-board on accessibility inspired by Kirsch.

Overall, the subject matter didactic school played a crucial role in challenging the mathematics education at the time, which was highly conceptually difficult and had many similarities to the practices in university mathematics [Hofe and Blum, 2016]. Today the scene of mathematics education is completely different than in the 60s, however the subject matter didactic analysis, has kept its relevance. This type of analysis works through mathematical or statistical content preparing it for students or teachers to read [Vohns, 2016]. A similar

’content-orientated’ analysis can be found in the French didactic tradition. Namely the a priori analysis within the school of didactic engineering [Hußmann et al., 2016]. I have chosen to use both analyses types in my preparatory work, as they despite similarities provide some very different perspectives. Now, I will explain the analysis types.

Subject Matter Didactic Analysis: A subject matter didactic analysis must process and conceptualize the mathematical content primarily with methods coming from the mathematical content itself. In short, the method consists of the teacher thoroughly and systematically studying the mathematical concept with the purpose of making it accessible for students or other teachers. The teacher acts as a kind of ”conveyor” of the content: organizing, reconstructing and declaring knowledge [Hußmann et al., 2016].

A Priori Analysis: This approach also work systematically with the content, but has a more student-active and environment-situated view, where students competencies and learning outcomes are forecast to strengthen the analysis [Hußmann et al., 2016]. The French approach, a priori analysis, often includes: a forecast of the target group of students, their mathematical prerequisites, potential misconceptions and strategies [Hußmann et al., 2016].

The full methodology of didactic engineering is not used, as my thesis does not have a practical part: performing a teaching sequence and conducting a posteriori analysis. The a priori analysis in this thesis is used more as a complement to the subject matter didactic analysis. This is to include student-active and environment-situated considerations in developing the didactical design.

2.4 Fundamental Ideas and Grundvorstellungen

In this thesis the two principles: ”fundamental ideas” and ”Grundvorstellung” are elaborated. These principles relates to a subject matter didactic way of analyzing content [Scheiner et al., 2023]. *Fundamental ideas* is the underlying ideas and methods that connect different parts of mathematics. Identifying fundamental ideas helps us provide a broader and more interconnected perspective on the subject of mathematics [Vohns, 2016]. Hence, fundamental ideas are main principles or the essence of mathematics and can, structurally, be regarded as global or overarching ideas [Scheiner et al., 2023]. In table 1 some examples of fundamental ideas are listed. Students often do not encounter fundamental ideas as inherent to a single topic, but it is something they acquire over time by identifying recurring features of mathematics [Vohns, 2016].

Examples of fundamental ideas in mathematics of Schreiber (1979)
Properties: quantity, continuity, optimality, invariance and infiniteness
Concepts: ideation, abstraction, representation, space and unit

Table 1: Examples of fundamental ideas [Scheiner et al., 2023]

The principle of *'Grundvorstellungen'* comes from the subject matter didactic view of mathematics focusing on how one could simplify mathematics to match with peoples cognitive abilities and human experiences [Hofe and Blum, 2016]. The application of the concept comes from the German mathematics educator Wilhelm Oehl (1904-1991). According to Oehl, Grundvorstellung refers to the idea or meaning behind a specific aspect of mathematics such as methods, properties or mathematical operations [Hofe and Blum, 2016].

Aspect: A subdomain of a concept that can be used to characterize the concept.
Grundvorstellungen: A conceptual interpretation that gives meaning to the aspect.

Table 2: "Aspect–Grundvorstellung" relation [Greefrath et al., 2016]

A description of aspects and Grundvorstellungen is given in table 2 above. Connected to one aspect there can be different Grundvorstellung, i.e., ways of understanding the mathematical aspect, such as through various verbal and graphical representations. Thus, it makes sense to work in a dynamic way, constantly re-representing mathematical aspects in different ways and formulating different Grundvorstellungen and layers of understanding [Hofe and Blum, 2016]. The role of subject matter analysis is among others to establish relationships "Aspect–Grundvorstellung" of a given mathematical concept — and to support the students "process of concept formation" [Greefrath et al., 2016]. In my thesis, I identify fundamental ideas, aspects and Grundvorstellungen in the subject matter didactic analysis of MLE, examining layers of meaning and establish relationships between likelihood functions, MLE and concepts in school mathematics.

2.5 Statistics Education

Since I have found almost no subject-didactic literature on teaching MLE aimed at upper secondary school, I have drawn upon subject-didactic research on statistics education more generally. Statistics education refers to the academic body of scholarly work that focuses on how statistics is taught and learned [Zieffler et al., 2017]. I will not go into detail about the field of research, as a wide range of research questions are being explored. However, I will briefly outline some of the concepts from the literature that I draw upon in the two analyses.

— **Statistical and mathematical reasoning:** Statistics and mathematics are different in their nature of reasoning. *Mathematical reasoning* is logical and deductive and involves applying axioms and definitions, and finding logical patterns by use of configuration and abstraction [Ottaviani, 2011]. *Statistical reasoning* is also about finding patterns, but it differs a lot as it rely on data. Statistical reasoning is inductive: it is about formulating questions, collecting, analyzing and interpreting data [Burrill and Biehler, 2011].

— **Statistical literacy:** Students ability to understand statistical terminology in a context and to understand argumentation based on statistics [Burrill and Biehler, 2011]. This is also the ability to reason statistically, apply statistical knowledge and to develop a critical sense in today's information society [Batanero and Borovcnik, 2016].

— **Statistics-as-magic:** The term ”statistics-as-magic” is used by G. W. Cobb and D. S. Moore in [Cobb and Moore, 1997]. The term refers to the fact that students have no real understanding of what is happening in a statistics, because statistical knowledge is taught in a way that relies heavily on digital tools and automated recipes. The digital tools perform all the hard work for the students [Cobb and Moore, 1997]. This term is similar to what [Pedersen and Jankvist, 2021] mention as black-box using CAS-tools.

— **Shared problem space:** Tasks and concepts that build a bridge between mathematics and statistics [Groth, 2015].

3 Research Questions

Based on the theoretical framework, the goal of the my investigation is to build a strong foundation for the didactical design and thereby answer the main research question of this thesis. This thesis is divided into two main sections: a subject-matter didactic analysis (in section 6) and an analysis of the didactical design (in section 7). The subject matter didactic analysis lead to two following areas of investigation and related research questions:

1. The concept formation of MLE approached by subject matter didactic.

   1.a) What knowledge will students acquire when working with the concepts of statistical model, likelihood function and MLE?

2. Legitimization of MLE as knowledge to be taught approached by subject matter didactic and research in statistics education.

   2.a) How can likelihood functions and MLE strengthen the content in statistics in upper secondary school?

   2.b) How can likelihood functions and MLE establish connections between different branches of school mathematics in upper secondary school?

The didactical design lead to following areas of investigation and related research questions:

3. An institutional analysis of the textbook chapter carried out by ATD.

   3.a) What institutional constrains and conditions are encountered when you design a textbook chapter about likelihood functions and MLE?

4. An a priori analysis comprised by praxeological organization and task design approached by ATD and research in statistical education.

   4.a) How can a teaching material on MLE be organized in a textbook chapter?

   4.b) How can tasks be formulated about MLE in a way that aligns with Danish upper secondary students' prerequisites and academic level?

4 History of Likelihood

In this section, I want to elaborate on the history of likelihood functions and MLE before conducting the subject matter didactic analysis. This is to help the reader better grasp these concepts. Many historical detail about likelihood are given in A. W. F. Edwards' article *"The History of Likelihood"* (1974) and in the book *"In all likelihood: Statistical Modelling and Inference Using Likelihood"* (2001) by Yudi Pawitan. I will skip most details and only give a rough outline beginning with the invention of probability and statistics.

Probability was invented by the French mathematicians Blaise Pascal (1623–1662) and Pierre de Fermat (1601–1665) in the 17th century and was motivated by their interest in gambling. The field of statistics was first born a century later as "probabilistic inference" by Thomas Bayes (1702–1761) and Pierre-Simon Laplace (1749–1827) [Batanero and Borovcnik, 2016].

Bayes and Laplace invented "probabilistic inference" independently of each other by introducing the concept of *Inverse Probability*. Today the term inverse probability is not used, but it is simply the term for a conditional probability of a hypothesis (H) given a sample ($Data$), denoted as $P(H|Data)$. We also know inverse probability from Bayes' theorem as the posterior distribution [Pawitan, 2001c]. It should be interpreted as the probability of a hypothesis being true, conditioned on observed data. To use Bayes' theorem, we assume that we know the probability of observing the data given the hypothesis, *a likelihood*, and that we know the underlying distribution of the hypothesis, *the axiomatic prior* [Pawitan, 2001c]. Then using Bayes' theorem, we can calculate the inverse probability as

$$\underbrace{P(H|Data)}_{\text{The inverse prob.}} = \frac{\underbrace{P(Data|H)}_{\text{The likelihood}} \times \underbrace{P(H)}_{\text{Axiomatic prior}}}{\underbrace{P(Data)}_{\text{Normalizing constant}}}.$$

Bayes' theorem includes determining the sampling distribution aka. the likelihood after data is observed. The concept of likelihood is fundamental in Bayes theorem as it contains our observed knowledge [Edwards, 1974]. In Bayes' theorem, be aware that the likelihood is combined with a prior, which means that statistical evidence is not obtained from the likelihood itself [Pawitan, 2001c]. Another important contribution to statistics was Carl Friedrich Gauss' (1777–1855) theory of errors, including the method of ordinary least squares [Batanero and Borovcnik, 2016]. Statistics as a systematic science that draws inference and is used for decision-making is relatively new. The rigorous mathematical approach used in statistics today, including the formal and systematic way of conducting statistical tests, is often referred to as modern statistics. In fact, statistics as a modern science is the newest branch of school mathematics [Varberg, 1963]. Modern statistics was pioneered by Ronald A. Fisher, Jerzy Neyman, and Karl Pearson from the 1920s to the 1940s, where they systematized confidence intervals and statistical tests [Batanero and Borovcnik, 2016].

One of the most important concepts of this thesis is the notion of likelihood, which like the method of MLE was invented by Fisher between 1912 and 1922 [Edwards, 1974]. It is essential to distinguish Fisher's notion of likelihood from Bayes' theorem, where $P(Data|H)$ is used to describe the conditional probability of drawing a sample given that a hypothesis is true. In Fisher's notion, the likelihood function can be used to gain statistical evidence solely based on the observed data and the function itself [Etz, 2018]. I will denote a likelihood as $L_{Data}(H)$ and it is a function defined in terms of the conditional probability

$$L_{Data}(H) = P(Data|H),$$

The likelihood function measures the relative possibility of the occurrence of the observed sample ($Data$) given specific choices of hypothesis (H) [Reid, 2000].

The difference between a likelihood function and a conditional probability is which one of ($Data$) and (H) is considered varying and which one is fixed. For a conditional probability the sample is considered varying and the hypothesis is considered fixed. On the contrary, a likelihood function is varying over possible hypothesis and the observed sample is considered fixed [Etz, 2018]. MLE is the method of finding the hypothesis that best explain the observed sample and it is found by maximizing the likelihood function. We will soon return to these concepts and explain them in more detail in the subject matter didactic analysis.

5 Brief Introduction to Probability

I will now introduce the basics of probability theory as a foundation for presenting key concepts such as statistical models and likelihood functions.

Random experiment	An *experiment* where the outcome cannot be predicted with certainty but if it is repeated many times, certain 'probabilistic' patterns can be observed.
Sample space $\mathcal{X}$	The *sample space* $\mathcal{X}$ of an experiment is the set of all possible outcomes of an experiment.
Event A	An *event* A is the subset of the sample space $\mathcal{X}$.
σ-algebra Measurable space $(\mathcal{X}, \mathcal{A})$	A *σ-algebra* is a family of subsets of $\mathcal{X}$ that obey certain axioms. We require that the set of events $\mathcal{A}$ is a σ-algebra in order to assign events with probabilities. If $\mathcal{A}$ is a σ-algebra on $\mathcal{X}$, we call the pair $(\mathcal{X}, \mathcal{A})$ a *measurable space*. We will mainly use two types of measurable spaces. The type where $\mathcal{X}$ is a countable set and $\mathcal{A} = \mathbb{P}(\mathcal{X})$ is the power set, and the type where $\mathcal{X} = \mathbb{R}^n$ for some n and $\mathcal{A} = \mathbb{B}(\mathbb{R}^n)$ is the Borel-σ-algebra on $\mathbb{R}^n$.
Probability measure P Probability space $(\mathcal{X}, \mathcal{A}, P)$	A *probability measure* on a measurable space $(\mathcal{X}, \mathcal{A})$ is a function $P : \mathcal{A} \to [0, 1]$ that satisfies $P(\mathcal{X}) = 1$ and for every countable sequence $(A_n)_{n \geq 1}$ of pairwise disjoint elements of $\mathcal{A}$ it hold $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$. We call $(\mathcal{X}, \mathcal{A}, P)$ a *probability space*.
Random variable X	Given a probability space $(\mathcal{X}, \mathcal{A}, P)$ and another measurable space $(E, \mathbb{E})$ a *random variable* is a measurable function $$X : \mathcal{X} \to E.$$

Probability distribution	Let $X : (\mathcal{X}, \mathcal{A}, P) \to (E, \mathcal{E})$ be a random variable. Then the *probability distribution* of X is the map $P^X : \mathcal{E} \to [0,1]$ given by $$P^X(B) = P(X^{-1}(B)) \qquad \forall B \in \mathcal{E}.$$
Density function	Given a random variable $X : (\mathcal{X}, \mathcal{A}, P) \to (E, \mathcal{E})$ and a measure Q on $(E, \mathcal{E})$, we say that X has a *density function* with respect to Q if there exists a measurable function $f : E \to \mathbb{R}_+$ such that $$P^X(B) = \int_B f \, dQ, \quad \forall B \in \mathcal{E}.$$ We will only consider two types of density functions. If $\mathcal{X}$ is countable and $\mathcal{A} = \mathbb{P}(\mathcal{X})$, then we only consider densities where Q is the counting measure. In this case the function $f : E \to \mathbb{R}_+$ is called the point mass probability. If $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{A} = \mathbb{B}(\mathbb{R}^n)$, then we only consider densities where Q is the Lebesgue measure on $\mathbb{R}^n$ and then the integral becomes the integral we know from calculus.

The references of this list are "Probability Essentials" of [Jacod and Protter, 2012] and "Measures, Integrals and Martingales" of [Schilling, 2017].

This list can be used as a reference for the later sections, but having a deep understand of all the concepts is not necessary. We will generally not discuss the framework of σ-algebras further, but simply assume that things behave nicely and are measurable. This decision comes from the fact that this construction is not core content in Danish upper secondary schools [Børne- og Undervisningsministeriet, 2024a].

Often, it is also skipped in introductory university statistics courses, since the probabilistic foundation is less prominent — and we often make nice choices such as $\mathcal{A} = \mathbb{P}(\mathcal{X}^n)$ or $\mathcal{A} = \mathbb{B}(\mathbb{R}^n)$ for the two cases [Lauritzen, 2023].

6 Subject Matter Didactic Analysis

In this section, I will introduce the terminology and statistical theory behind likelihood functions and MLE and conduct a subject matter didactic analysis. The concept formation, is heavily inspired by the article *“Aspects and ‘Grundvorstellungen’ of the Concepts of Derivative and Integral: Subject Matter-related Didactical Perspectives of Concept Formation”* by G. Greefrath, R. Oldenburg, H.-S. Siller, V. Ulm, and H.-G. Wiegand [Greefrath et al., 2016]. My analysis draws on the article’s use of the analytical tools ’aspect’ and ’Grundvorstellung’, but I also use the article’s structure to organize my own investigation.

Although the subject matter didactic analysis focuses on the subject matter (the statistical content of MLE), I occasionally include input related to the target group, level of complexity and the core content of the syllabus in upper secondary school. This was necessary to set boundaries for the investigation of MLE, which could otherwise have been much more extensive and more theoretical without these limitations.

6.1 The ”Statistical Model” Concept

First, I will elaborate on the ”statistical model” concept, which may be the most important underlying feature of likelihood functions and MLE. I begin by clarifying how the ”statistical model” concept is approached in this thesis and I investigate how it is incorporated in Danish upper secondary school textbooks.

The German mathematicians Uwe-Peter Tietze, Manfred Klika, and Hans Wolpers describe in their book *”Mathematikunterricht in der Sekundarstufe: Didaktik der Stochastik”* that, statistical modelling involves the description of stochastic situations and the formalization of these situations. This is done by the use of specific probability distributions such as the binomial distribution and the normal distribution [Tietze et al., 2002]. Furthermore, they characterize the ”statistical model” concept as a mathematical structure given by $\mathcal{X}$, where $\mathcal{X}$ is a sample space, and a family $(P_\theta)_{\theta\in\Theta}$ of probability distributions [Tietze et al., 2002]. This structure is very general and makes it possible to describe many different stochastic situations from real-world examples to artificial situations with dice, playing cards and coin tosses.

It is the specific choices of $\mathcal{X}$ and $(P_\theta)_{\theta\in\Theta}$, made appropriately in relation to a stochastic situation, that constitute the model [Tietze et al., 2002]. This rather strict characterization of a statistical model is in this thesis regarded as the ”parameterized family” aspect and is used later in *Definition 1*. Consistent with this view, the ”statistical model” concept is regarded as a mathematical description of a stochastic situation. The subject-didactic analysis focuses on the conceptualization of this, asking what characterizes a statistical model and how a statistical model obtains its meaning.

I have investigated how statistical models are integrated into Danish upper secondary school mathematics education. This has not been straightforward to figure out. The investigation was conducted based on the syllabus and a selection of textbooks, in which I searched for the terms ”statistical model” and ”binomial model” and examined their subject-specific context. The following material was included in the investigation.

— Matemat10k : statistik [Agermose Jensen and Timm, 2014].

— Lærebog i matematik A2 stx [Brydensholt and Ebbesen, 2025a].

— Lærebog i matematik A3 stx [Brydensholt and Ebbesen, 2025b].

— Mat A2 stx [Carstensen, 2021].

— Sandsynlighedsregning og statistik [Clausen et al., 1997].

— Hvad er matematik? – B: grundbog [Grøn, 2015].

— Statistik C [Gråbæk et al., 2025].

— Højniveaumatematik. Bind 2 [Hebsgaard and Sloth, 1999].

— Læreplan matematik A – stx [Børne- og Undervisningsministeriet, 2024a].

— Vejledning til læreplan i matematik A [Børne- og Undervisningsministeriet, 2024b].

We will refer to this list as Material List. Among the textbooks I have had access to, none include a formal definition of a statistical model. Several of the textbooks do not use the term ”statistical model” at all, although some do provide a formal definition of the binomial model [Clausen et al., 1997]. The concept of ”mathematical model” is mentioned in the syllabus for upper secondary school (A-level), but the term ”statistical model” is not [Børne- og Undervisningsministeriet, 2024a]. On this basis, the material suggest that statistical models are either included in the form of the binomial model, or not included.

Likewise, the investigation shows that the ”parameterized family” aspect is not included in the mathematics textbooks. Instead my investigation shows, that particular examples play a key role in teaching statistical models. Also Tietze, Klika and Wolpers emphasize that especially artificial situations (examples such as games involving dice, playing cards, or coin tosses) are the common way of teaching the ”statistical model” concept. These types of situations typically give rise to a binomial model, and the preferred approach is to teach the concrete binomial model rather than the general concept of statistical models [Tietze et al., 2002]. In terms of the subject matter didactic framework, I think the binomial model and the mentioned artificial situations are important as applications to statistical models and helps establish intuition. In contrast, I look at the general concept of a statistical model.

6.1.1 The ”Parameterized Family” Aspect

A statistical model relies as mentioned on a stochastic situation. This is represented by one or more data observations $x = (x_1, x_2, ..., x_n)$. We formalize that data is stochastic as a realization of a random variable $X = (X_1, X_2, ..., X_n)$. By the ”parameterized family” aspect a statistical model is characterized by defining the sample space $\mathcal{X}$ and by indicating the distributions that can reasonably be assumed to have generated the given data as a family of distributions [Ditlevsen and Sørensen, 2018]. If we disregard the σ-algebra, we can summarize a statistical model in three components based on the ”parameterized family” aspect. After this, a more technical definition is given also based on the ”parameterized family” aspect now including the σ-algebra.

Three Components of a Statistical Model

1) The sample space $\mathcal{X}$ of an experiment is the set of all possible outcomes of an experiment.

2) An observation $x = (x_1, ...x_n)$ that represents the occurrence of one or more experiment.

3) A family of possible probability distributions $(P_\theta)_{\theta\in\Theta}$ on the sample space $\mathcal{X}$.

[Ditlevsen and Sørensen, 2018]

Definition 1. A statistical model contains a measure space $(\mathcal{X}, \mathcal{A})$ and a family $(P_\theta)_{\theta\in\Theta}$ of probability distributions on $(\mathcal{X}, \mathcal{A})$. The space $(\mathcal{X}, \mathcal{A})$ is called the representation space, where $\mathcal{X}$ is the sample space and $\mathcal{A}$ is a σ-algebra. The parameter space Θ is the index set for the family of probability distributions in the model.

Further, we assume that we have a family of densities $(f_\theta)_{\theta\in\Theta}$, where f_θ is a density function for P_θ [Lauritzen, 2023].

Remark: In the case where the sample space $\mathcal{X}$ is discrete we call f_θ a point mass function.

In this definition I have made several indirect assumptions. When writing $(P_\theta)_{\theta\in\Theta}$, I have assumed that the family of probability measures is *parameterized*, i.e., the family can be indexed by a parameter space Θ. I have also assumed that every measure has a density, in the general setting we would say the family is *Dominated*. A density is always given with respect to another measure, and I will always use the Lebesgue measure if the data is continuous and the counting measure if the data is discrete [Lauritzen, 2023]. The assumptions that the family is parameterized and dominated is necessary in order to define the likelihood function later. By this aspect we obtain the following characterization of the binomial model.

Let X be a random variable on the sample space $\mathcal{X} = \{0, 1, 2, ..., n\}$ and $\Theta = [0, 1]$ be a parameter space. Let $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ be a family of probability distributions given by

$$P_\theta(X = x) = \binom{n}{x}\theta^k(1-\theta)^{n-x}, \qquad x \in \{0, 1, 2, ..., n\}.$$

Then P_θ is the binomial distribution and the above equation is the point mass function. This is a discrete statistical model specified by the binomial distribution [Lauritzen, 2023].

6.1.2 The "Random Variable" Aspect

An alternative aspect is the "random variable" aspect. In this aspect, I view a statistical model as a random variable with a distribution including an unknown parameter. In the case of the binomial model, I may write: Let X be a stochastic variable with sample space $\mathcal{X} = \{0, 1, .., n\}$ and suppose that $X \sim bin(n, \theta)$ where θ is an unknown parameter in $\Theta = [0, 1]$ [Ditlevsen and Sørensen, 2018].

The important difference between this aspect and the "parameterized family" aspect is whether I consider one distribution with a varying parameter θ or a set of many distributions. The two aspects are very similar, but the "random variable" aspect is more suitable for an example like tossing a coin. Since the focus is on the random variable, which represent the outcome of the coin. The "random variable" aspect is also the chosen aspect in the school material that define a binomial model. This aspect is used in [Clausen et al., 1997].

6.1.3 The Grundvorstellung "Prior Assumptions"

A statistical model gains its meaning within the reality or the context associated with the stochastic situation. The context of the stochastic situation reflects various features and conditions, which one in statistics refers to as prior assumptions [Milhøj, 2025]. This gives rise to the Grundvorstellung I call "prior assumptions". For example, the interpretation of a binomial model arises from the specific assumptions underlying a binomial experiment. Namely, that there are two possible outcomes, that we are considering repeated trials, and that the outcomes are independent of one another [Clausen et al., 1997].

In general, one can divide prior assumptions into two types: the structural part and the stochastic part [Milhøj, 2025]. The structural part of the binomial model consists of the features we just described for a binomial experiment. For instance, in the case of coin tossing, the interpretation of the statistical model relies on the repetition of an experiment with the two possible outcomes heads and tails. If the condition is that the coin is tossed 10 times and we count the the outcome of heads, the sample space becomes $\mathcal{X} = \{0, 1, \ldots, 10\}$. These prior assumptions form the basis for why the binomial distribution is the appropriate distribution for this type of stochastic situation.

It is primarily through this contextualization of these structural elements that the aspects of a statistical model gain its meaning. In the "parameterized family" aspect, the structural part is reflected in the specification of the sample space and in the family of distributions. Likewise, in the "random variable" aspect, it is reflected in the choice of the sample space and the assumption that the random variable X lies in a specific class of probability distributions.

The second part is the stochastic component, which is a major theme in the statistics education literature. It is referred to by various terms like statistical uncertainty, non-reproducibly or variability [Makar and Rubin, 2017]. A statistical model also obtains its meaning from the uncertainty associated with data. When drawing a sample, one only has access to partial information, and this gives rise to uncertainty [Ditlevsen and Sørensen, 2018]. This can be explained according to the terms of population, sample and non-reproducibly.

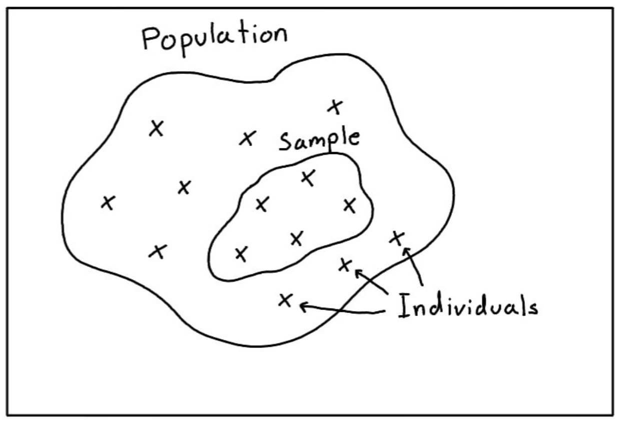


Figure 3: Sample, Individuals and Population [Appendix B].

If a sample is randomly chosen from all the possible samples that could have been drawn, then another randomly selected sample will likely not have precisely the same features as the first one. For example, if we toss a coin 10 times over many repetitions of the experiment, we will not get exactly five heads and five tails each time, since the experiment is random. So uncertainty means that when we draw a new sample, the outcome is not reproducible [Pawitan, 2001c]. The basic idea of drawing a sample is shown in figure 3.

Considering the aspects, the stochastic part in the ”parameterized family” aspect, is contained within the individual distributions of the family. Meaning that the individual distributions formalize the uncertainty connected to the non-reproducibility of the sample. In the ”random variable” aspect, the stochastic part is represented through the different choices of the parameter θ. I conclude, that the grundvorstellung ”prior assumptions” contains two parts, which both give meaning to the aspects. This connection is illustrated in figure 4.

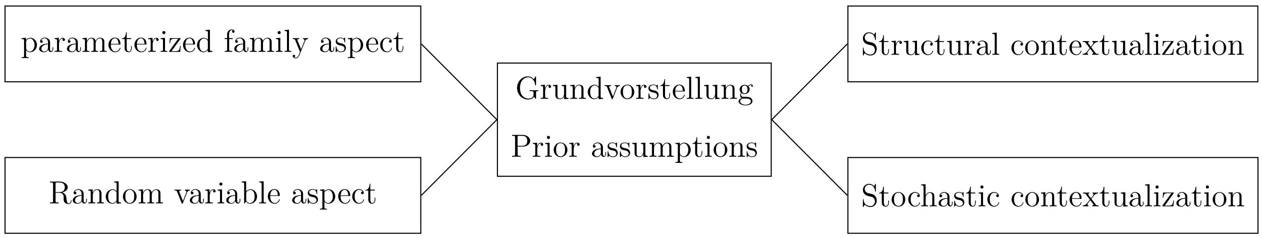


Figure 4: Aspect and grundvorstellung of statistical models.

According to my analysis, the Grundvorstellung ”prior assumption” can be transferred to the following competencies, that students can acquire.

- A1: Recognize the structural and the stochastic parts of stochastic situations. Especially focusing on identifying these two parts in relation to binomial experiments as the stochastic situation.
- A2: Explain how stochastic situations (either real-world situations or artificial situations) can be translated into mathematical symbolic language. This translation should be based on either the ”parameterized family” aspect or the ”random variable” aspect.
- A3: Can switch between the contextualized level and the mathematical symbolic level of the statistical model.

6.1.4 Introduction to iid. Random Variables

This introduction is a technical elaboration that will give me some tools. I will use these tools in the case where I have observed data $x = (x_1, ..., x_n)$ to make a statistical model of this situation.

Assumptions. For the random variable $X = (X_1, ..., X_n)$ with the outcome $x = (x_1, ..., x_n)$ I make the following three assumptions

1. The random variable X_i with $i \in \{1, ..., n\}$ is specified by a probability distribution or a probability density function.

$$f_{X_i}(x|\theta) \qquad i \in \{1, 2, ..., n\},$$

where x is a possible outcome and θ is the parameter specifying the density.

2. The random variables X_i and X_j with $i, j \in \{1, ..., n\}$ are identically distributed, i.e., $f_{X_i}(x|\theta) = f_{X_j}(x|\theta)$ for all i and j.

3. The random variables X_i and X_j with $i, j \in \{1, ..., n\}$ and $i \neq j$ are independent, i.e.,

$$f_{X_1,...,X_n}(x_1, ..., x_n|\theta) = f_{X_1}(x_1|\theta) \cdot ... \cdot f_{X_n}(x_n|\theta) = \prod_{i=1}^{n} f_{X_i}(x_i|\theta).$$

[Nielsen, 2017]

If I have n observations from a sample space $\mathcal{X}$ and I assume the observations to be iid., then we have the statistical model on the new sample space $\mathcal{X}^n$ with $\mathcal{P} = \{P_\theta^{\oplus n}|\theta \in \Theta\}$. Since we have iid. observations the densities become $f_\theta(x) = \prod_{i=1}^{n} f_\theta(x_i)$ [Lauritzen, 2023].

In the textbooks from the Material List, neither the assumption of iid. random variables or the fact that the joint density is the product of the marginal densities is mentioned. However, some materials does cover the summation symbol and basic rules for working with it [Gråbæk et al., 2025]. This is sometimes introduced in connection with numerical integration [Hebsgaard and Sloth, 1999]. The concept of the product symbol can be transferred from the summation symbol. Note that assumption 3 above can be seen as a generalization of the multiplication principle, where students learn that the probability of two independent events occurring can be calculated by multiplying the probabilities of each event: $P(\text{both A and B}) = P(A)P(B)$ [Brydensholt and Ebbesen, 2025a].

A normal model characterized by the "parameterized family" aspect and with the assumption of iid. random variables is given by the following description.

Let $X_1, ..., X_n$ be iid. random variables on $\mathcal{X} = \mathbb{R}$ and assume X_i is normal distributed with parameters μ and σ, i.e. they have densities

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The parameter space is $\Theta = \mathbb{R} \times \mathbb{R}_+$. Our new statistical model have representation space $(\mathbb{R}^n, \mathbb{B}(\mathbb{R})^n)$ and the parameter space is still $\Theta = \mathbb{R} \times \mathbb{R}_+$. The distribution are given as a product distribution with densities

$$f_{\mu,\sigma}(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

6.2 The "Likelihood Function" Concept

The "likelihood function" concept is fundamental in statistics and plays a central role in many statistical inference methods. The most well-known methods based on likelihood functions are MLE, likelihood-ratio test and Bayesian inference [Etz, 2018]. Such methods are commonly used in both theoretical statistics and applied statistics, particularly in fields like machine learning [Efron and Hastie, 2016] and psychology [Etz, 2018]. Despite the wide application of the concept, there are relatively few subject-didactic sources on the topic. A brief subject-didactic analysis of likelihood functions can be found in German by D.-P. Tietze, M. Klika, and H. Wolpers in *"Mathematikunterricht in der Sekundarstufe: Didaktik der Stochastik"* [Tietze et al., 2002]. A somewhat more in-depth subject-didactic analysis can be found in Alexander Etz's article *"Introduction to the Concept of Likelihood and Its Applications"* [Etz, 2018]. This article has its primarily target group as psychology students at the universities and not upper secondary school teachers.

The subject-didactic analysis of the "likelihood function" concept in this thesis is largely based on my own original analysis, since there are few subject-didactic sources available. On the other hand, there are many theoretical statistical sources that present the concept of likelihood functions. In this analysis, the source for the theoretical foundation have primary been: Steffen Lauritzen's book *"Fundamentals of Mathematical Statistics"* [Lauritzen, 2023], Susanne Ditlevsen and Helle Sørensen's lecture notes *"Introduktion til Statistik"* [Ditlevsen and Sørensen, 2018], and Heino Bohn Nielsen's book *"Introduction to Likelihood-based Estimation and Inference"* [Nielsen, 2017].

6.2.1 The Intuitive "Likelihood Function" Concept

Since likelihood functions are first taught at an academic level, I can not assume that upper secondary school students are familiar with the concept. Also, I cannot assume that upper secondary school mathematics teachers necessarily remember the topic from their time at university. I must therefore start from the basics by addressing two general questions:

— What is the setting where the "likelihood function" concept can be used?

— What is the purpose of introducing the "likelihood function" concept?

(1) The setting where the likelihood function can be used.

Likelihood functions apply to statistical sampling, when knowledge about the population is incomplete. The incompleteness of knowledge arises from, the fact that I have no prior information about the underlying distribution from which data is drawn. In such cases, our only source of information is the data itself. Thus, introducing likelihood functions consists of two main elements: (1) I have a sample, and (2) I assume that I have a family of distributions (also known as a statistical model) based on that sample [Etz, 2018]. This setup has been explained in detail in the previous section on statistical models.

(2) The purpose of introducing likelihood functions.

The purpose of the likelihood function is to gain information about an unknown quantity. Within a statistical framework, this unknown quantity is represented by a parameter θ in a statistical model. Sometimes, the purpose is to estimate the unknown parameter θ, but I shall show later that other types of analysis of θ can also be gained from likelihood functions [Etz, 2018].

In addition to these two general questions, the intuition behind likelihood can also be developed through examples. I will primarily use the binomial model and coin tossing as examples, with only a few exceptions. The reason for choosing the binomial distribution is that it is included in the upper secondary school mathematics syllabus and it is more simple than the normal distribution, which is also included in the A-level syllabus [Børne- og Undervisningsministeriet, 2024a].

Example 1: Coin tosses

To get an intuitive idea of the "likelihood function" concept, I will look at a classic setting with coin tosses. In this coin toss scenario, I have no prior knowledge about the distribution of heads and tails. Hence, I am uncertain whether the coin is fair. Let X be a random variable that indicates the number of heads in ten independent trials. After tossing the coin ten times, I observe $x = 9$ heads. Given the knowledge of this observation alone, what would the probability θ be of getting heads? The information I have about θ is incomplete, and I cannot say anything with certainty about the distribution of heads and tails. However, I can still have some ideas about the likelihood of heads and tails based on our experiment.

I might suspect that the coin is not fair and that the probability of head is high. I could estimate θ by guessing such as $\hat{\theta}_1 = 0.77$ or $\hat{\theta}_2 = 0.92$. The likelihood function provides us with a deductive way to compare different values of θ [Etz, 2018]. This is a function of the unknown parameter θ derived from the point mass probability

$$L(\theta) = P_\theta(X = 9) = \binom{10}{9}\theta^9 \cdot (1-\theta),$$

which is called the likelihood function. Then $L(\hat{\theta}_1)$ and $L(\hat{\theta}_2)$ are calculated as

$$L(0.77) = \binom{10}{9} \cdot 0.77^9 \cdot (1-0.77) = 0.219,$$

$$L(0.92) = \binom{10}{9} \cdot 0.92^9 \cdot (1-0.92) = 0.378.$$

Based on the sample, I can use the likelihood function to rank all values of θ according to how likely they make the sample [Fisher, 1922]. Since, it holds that $L(\hat{\theta}_2) > L(\hat{\theta}_1)$, I have that the point mass probability $P_{\hat{\theta}_2}$ with $\hat{\theta}_2 = 0.92$ explain the observation $x = 9$ better than the point mass probability $P_{\hat{\theta}_1}$ with $\hat{\theta}_1 = 0.77$.
I still do not know the true value of the parameter θ, since the likelihood function reflects the uncertainty within the statistical framework [Pawitan, 2001b]. However, the likelihood function is a rational-deductive measure that can provide the estimate that makes the observation most likely. This estimate "the best estimate" is exactly the MLE. MLE is calculated from maximizing the likelihood function [Fisher, 1922], as I will show later.

I will now summarize the intuitive "likelihood function" concept.

- B1: Perhaps the most central idea is the intuition of 'comparability' or 'ranking'. This is the intuition that the likelihood function based on data, provides a way to rank and compare different values of θ.

- B2: Another intuition is 'weakness' or 'incompleteness'. It is the intuition that likelihood functions can be applied to situations in which the knowledge of a stochastic situation is incomplete, and one therefore cannot make definitive conclusions.

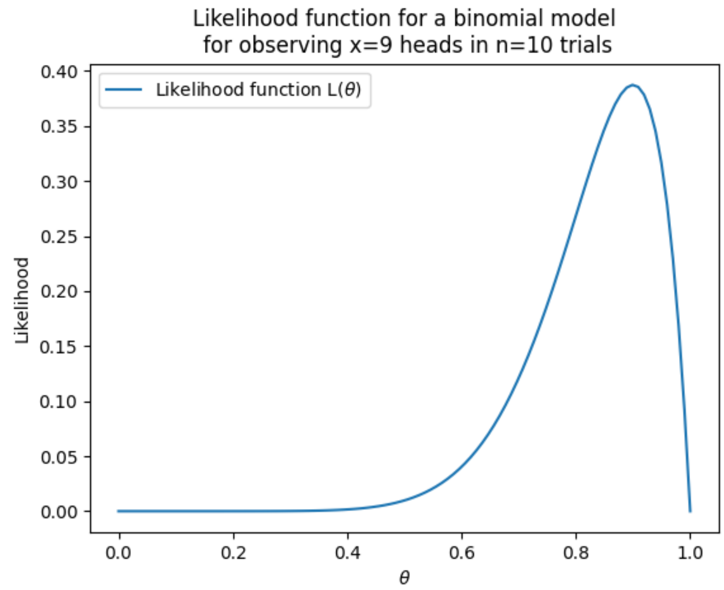


Figure 5: Likelihood function of the probability θ in a binomial model with $x = 9$ and 10 trials. I can observe graphically from the plot that $\hat{\theta}_2 = 0.92$ has a larger likelihood than $\hat{\theta}_1 = 0.77$ since $L(\hat{\theta}_2) > L(\hat{\theta}_1)$. The figure shows that the MLE is $\hat{\theta}_{MLE} = 0.9$, this is the frequency $\left(\frac{x}{n} = 0.9\right)$, which is quite intuitive [Appendix B].

6.2.2 The "Density" Aspect

In the following, I analyse the "likelihood function" concept in terms of subject matter didactic. I do so by establishing the relationship between aspects and Grundvorstellungen of the concept. In my analysis, I have only found one aspect of the "likelihood function" concept, namely the that the it is derived from the density $f_\theta(x)$ in the continuous case and the mass point probability $P_\theta(X = x)$ in the discrete case. In short, the "density" aspect is to view the density function as a function of the parameter θ, which is a precise characterization of the likelihood function [Fisher, 1922].

In *definition 2* a definition of the likelihood function is given, building on this aspect. In the definition I also define the log-likelihood function, which is the composition of the likelihood function as the inner function and the natural logarithm as the outer function. I will return later to why the log-likelihood function is particularly useful.

I introduce the likelihood function based on that I have observed some data $x = (x_1, ...x_n)$ and considering a statistical model with sample space $\mathcal{X}$, parameter space Θ and densities $(f_\theta)_{\theta\in\Theta}$. Then I can define the likelihood function and the log-likelihood function.

Definition 2. For every $x \in \mathcal{X}$, I define the likelihood function $L_x : \Theta \to [0, \infty)$ by

$$L_x(\theta) = f_\theta(x).$$

The log-likelihood $\ell_x : \Theta \to [-\infty, \infty)$ is found by taking the logarithm of the likelihood function

$$\ell_x(\theta) = \log(L_x(\theta))$$

[Lauritzen, 2023].

In the example with coin tosses in 10 trials, formalized by a binomial model, the likelihood function is characterized by the point mass probability of the binomial distribution, viewed as a function of the parameter θ. This function is given by $L_x(\theta) = \binom{10}{x}\theta^x \cdot (1-\theta)^{10-x}$.

Technicalities: In the context where I have observed n iid. observations $x = (x_1, ..., x_n)$ the likelihood function becomes $\prod_{i=1}^n L_{x_i}(\theta)$ and therefore the log likelihood function becomes

$$\ell_x(\theta) = \log\left(\prod_{i=1}^n L_{x_i}(\theta)\right) = \sum_{i=1}^n \log\left(L_{x_i}(\theta)\right) = \sum_{i=1}^n \ell_{x_i}(\theta). \tag{6.1}$$

6.2.3 The Grundvorstellung "Reversed Density"

The interpretation of the likelihood function as a "reversion" of the density is based on the swapping of x and θ in relation to which is fixed and which is varying.

$$L_x(\theta) = f_\theta(x)$$

This should be understood as considering the density function $f_\theta(x)$, as a function of two variables $f(x, \theta)$. The density function is given by considering the parameter θ as fixed and varying the function over x. For the likelihood function, it is the other way around, I now consider x as fixed and vary over the parameter θ instead. I call this conceptual understanding of the mathematical object the Grundvorstellung of "reversed density". Be aware that the likelihood function is not a density function over θ (or a probability mass function over θ in the discrete case), since it does not always integrate to 1. The likelihood function describes how possible/likely the observed data is given a specific parameter θ not how probable the parameter θ is, since the laws of probabilities do not hold for likelihoods [Fisher, 1922]. In the ongoing example with coin tosses for 10 trials and $x = 9$ the likelihood function is

$$L_x(\theta) = 10\theta^9(1-\theta).$$

I get the following integral when I integrate over the parameter space $\Theta = [0, 1]$.

$$\int_0^1 10\theta^9(1-\theta)d\theta = \int_0^1 \left(10\theta^9 - 10\theta^{10}\right) d\theta = \left[\theta^{10}\right]_0^1 - \left[\frac{10\theta^{11}}{11}\right]_0^1 = 1 - \frac{10}{11} = \frac{1}{11}. \tag{6.2}$$

So this example clearly does not integrate to 1.

So the Grundvorstellung "reversed density" has two important subject-didactic perspectives, that students can acquire.

- C1: First, the interpretation of the likelihood function by considering the density function as a function of the parameter θ for fixed x.
- C2: Secondly, a likelihood is not a probability density function or a point mass probability. This can be shown by an example as above in (6.2).

6.2.4 The Grundvorstellung "Relative Measure"

The Grundvorstellung "relative measure" is an extension to the intuition of 'comparability' or 'ranking'. As mentioned, the likelihood function is a way to rank specific values of the parameter θ by which one makes the observed data most likely. In this sense the likelihood function is a relative measure of how much a point estimate $\hat{\theta}$ supports the observed data. Suppose I have observed some data x and $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimates such that the probability distributions $P_{\hat{\theta}_1}$ and $P_{\hat{\theta}_2}$ are two possible candidates for the true data-generating distribution. If I have shown

$$L_x(\hat{\theta}_1) > L_x(\hat{\theta}_2), \tag{6.3}$$

then it is more likely that the data x was generated by $P_{\hat{\theta}_1}$ rather than $P_{\hat{\theta}_2}$. The intuition is that high outputs of the likelihood function and the log-likelihood function indicates that the data supports a probability distribution $P_{\hat{\theta}}$ as being the true data-generating distribution [Lauritzen, 2023].

The value of the likelihood function itself does not indicate how probable a specific value of θ makes data, because as mentioned the likelihood function is not a probability distribution [Etz, 2018]. So it is only meaningful to compare the likelihood of two or more different values of θ like I did in (6.3). Another way of doing it is by looking at the likelihood ratio

$$LR = \frac{L_x(\hat{\theta}_1)}{L_x(\hat{\theta}_2)}.$$

Returning to example 1 with coin tosses in 10 trials, I compared the specific values of θ, namely the estimates $\hat{\theta}_1 = 0.77$ and $\hat{\theta}_1 = 0.92$. The likelihood ratio for the two point estimates $\hat{\theta}_1 = 0.77$ and $\hat{\theta}_2 = 0.92$ is calculated by

$$LR = \frac{L_9(0.92)}{L_9(0.77)} = \frac{\binom{10}{9} \cdot 0.92^9 \cdot (1-0.92)}{\binom{10}{9} \cdot 0.77^9 \cdot (1-0.77)} = \frac{0.378}{0.219} = 1.726.$$

Therefore the observed data $x = 9$ is 1.726 times more probable under the hypothesis $\hat{\theta}_2 = 0.92$ than under the hypothesis $\hat{\theta}_1 = 0.77$ [Etz, 2018].

Another way of comparing the likelihoods is by visual inspection. Visual inspection allows us to see the full picture of all possible values of θ [Etz, 2018]. In Figure 6, the likelihood function is shown with dotted lines drawn at the estimated values. The lines are used for comparing the two estimates.

As mentioned, the likelihood function in itself has no interpretive meaning. Therefore, increasing the sample size n does not reflect the output of the likelihood function in any meaningful and interpretable way. However, increasing n is reflected in the likelihood ratio and in the visual inspection [Etz, 2018]. For example one can show that in the Grundvorstellung ”relative measure” it is of great importance whether I am observing 9 heads in 10 trials, 27 heads in 30 trials or 90 heads in 100 trials. It is intuitive that an experiment with 100 coin tosses resulting in 90 heads provides stronger statistical evidence than 10 coin tosses resulting in 9 heads, which is exactly what is reflected in the ”relative measure”.

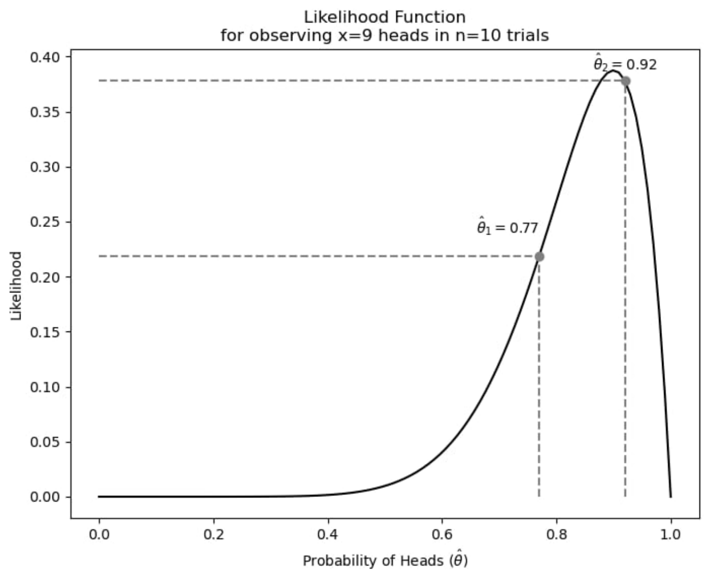


Figure 6: Visual inspection of the likelihood function with graph reading for $\hat{\theta}_1 = 0.77$ and $\hat{\theta}_1 = 0.92$ [Appendix B].

In relation to the visual inspection, the curve of the likelihood function becomes narrower as the number of trials (sample size) n increases. In Figure 7, I show the three mentioned cases with $n = 10$, $n = 30$, and $n = 100$ trials. It is clear from the figure that the ratio between the likelihood values for $\hat{\theta}_1$ and $\hat{\theta}_2$ is not the same across the three cases.

In relation to the likelihood ratio, I will now denote the likelihood functions by $L_{10}(\theta), L_{30}(\theta)$ and $L_{100}(\theta)$ respectively to distinguish between the three cases. Note that this is not entirely consistent with the notation above, since the subscript usually indicates the fixed observation x and not the sample size n.

In the case of 30 trials with $x = 27$, the likelihood ratio for the two competing hypotheses $\hat{\theta}_1 = 0.77$ and $\hat{\theta}_2 = 0.92$ is $\frac{L_{30}(0.92)}{L_{30}(0.77)} = 5.142$. This means that the observation is 5.142 times more likely under the hypothesis $\hat{\theta}_2 = 0.92$ than under the hypothesis $\hat{\theta}_1 = 0.77$. Comparing this to the likelihood ratio for 10 trials, the likelihood ratio for 30 trials is larger.

It makes sense that instead of comparing the two hypotheses $\hat{\theta}_1 = 0.77$ and $\hat{\theta}_2 = 0.92$, one should use the best-supported hypothesis ”the maximum likelihood estimate” as the reference value [Etz, 2018]. This gives the likelihood ratio $\frac{L(\theta)}{L(\hat{\theta}_{\mathrm{MLE}})}$.

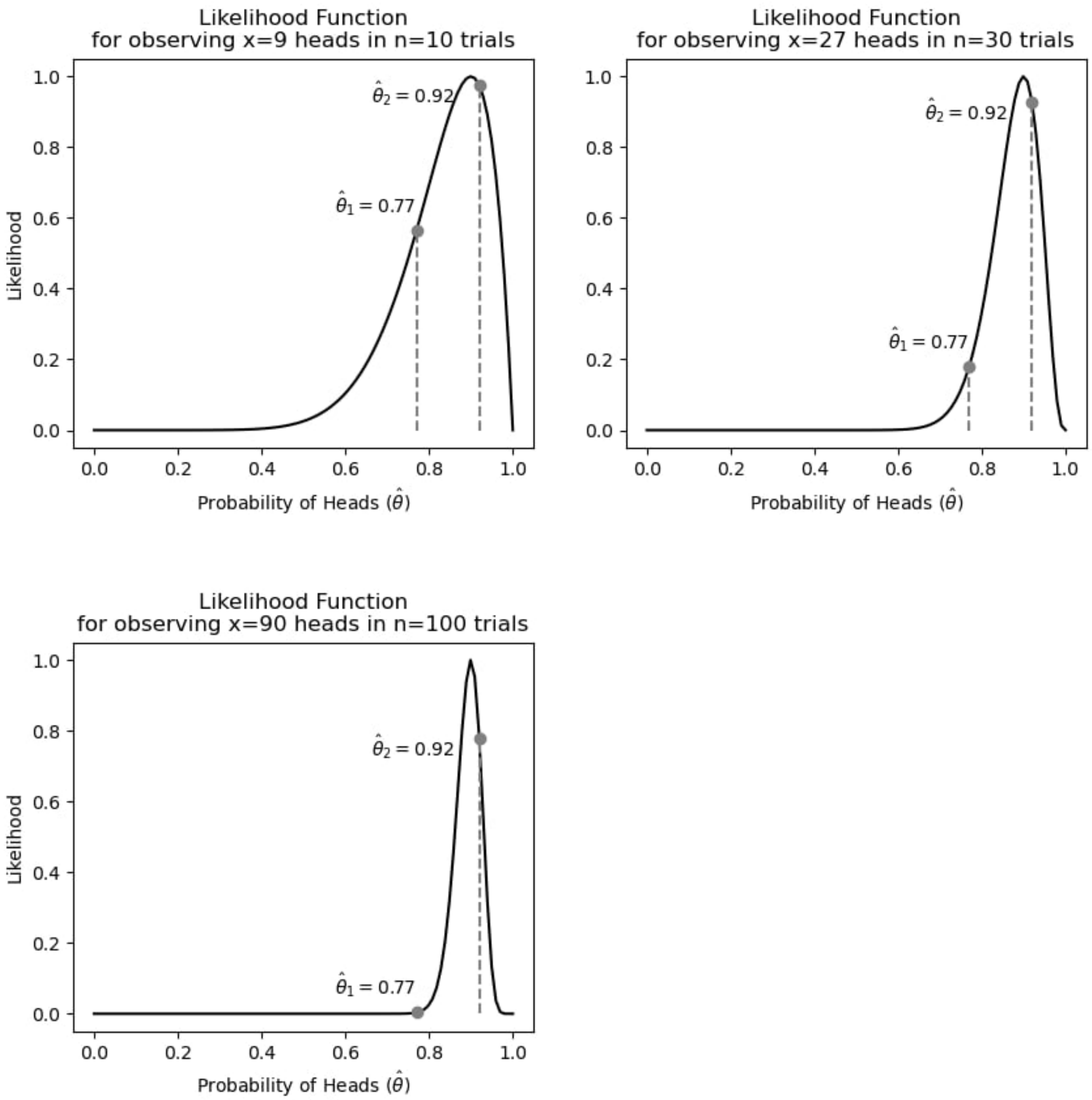


Figure 7: The scaled likelihood functions for observing $x = 9$ in 10 trials, $x = 27$ in 30 trials and $x = 90$ in 100 trials [Appendix B]. Scaled likelihood means that the likelihood function is scaled so it has maximum value corresponding to 1 (this does not chance the likelihood ratio).

When the hypothesis $\hat{\theta}_2 = 0.92$ is compared to the best-supported hypothesis $\hat{\theta}_{MLE} = 0.9$ in the three experiments, I obtain:

$$\frac{L_{10}(\hat{\theta}_2)}{L_{10}(\hat{\theta}_{MLE})} = 0.975 \qquad \frac{L_{30}(\hat{\theta}_2)}{L_{30}(\hat{\theta}_{MLE})} = 0.927 \qquad \frac{L_{100}(\hat{\theta}_2)}{L_{100}(\hat{\theta}_{MLE})} = 0.776$$

It is observed here that the statistical evidence becomes stronger for the best-supported hypothesis $\hat{\theta}_{MLE}$ compared to the hypothesis $\hat{\theta}_2$ as the number of observations n increases.

I conclude from this analysis that it is an important subject-didactic perspective to always look at the likelihood function in comparison. This is explained very clearly by Alexander Etz in *”Introduction to the Concept of Likelihood and Its Applications”*.

> *”We need to be careful not to make blanket statements about absolute support, such as claiming that the hypothesis with the greatest likelihood is “strongly supported by the data.” Always ask what the comparison is with.” [Etz, 2018]*

Even though the best supported hypothesis $\hat{\theta}_{MLE}$ has the highest likelihood of all the parameters, I see that in the case of 10 tosses the likelihood ratio of $\hat{\theta}_{MLE}$ and $\hat{\theta}_2$ is 0.975. So the observed data $x = 9$ has almost equal statistical evidence under the two hypothesis.
The grundvorstellung ”relative measure” can therefore develop students statistical literacy by making them deal with the reliability of the estimate.

The Grundvorstellung ”relative measure” can be summarized in the following terms that students can acquire:

- D1: Can describe the likelihood function as a relative measure of θ, in line with the intuition of comparison and ranking. This includes using the likelihood function to compare two hypotheses.

- D2: Use calculations of the likelihood ratio and visual inspection of the ratio to support statements about statistical evidence. These statement should explicitly mention the reference likelihood.

- D3: Be familiar with the impact that the sample size n has on the likelihood curve and on the likelihood ratio.

6.2.5 Summary of the ”Likelihood Function” Concept

From the analysis about the ”density” aspect, I establish the following.

- The likelihood function is characterized by the density or the point mass probability.

The Grundvorstellung refers to the interpretation that gives the aspects its meaning.

- The ”density” aspect gains meaning as a kind of ”reversed density”, where x and θ are swapped such that θ is varied, while x is fixed.

- The ”density” aspect also gains meaning as a ”relative measure”, allowing us to compare different values of θ based on how well they explain the observed data.

The first Grundvorstellung ”reversed density” emphasizes a semantic understanding of the concept, focusing on the likelihood function as a mathematical object. The latter Grundvorstellung ”relative measure” emphasizes a heuristic perspective, viewing likelihood as a tool in statistical inference.

6.3 The ”Maximum Likelihood Estimation” Concept

The hypothesis that best supports the observed data is, the maximum likelihood estimate. In, other words the MLE is the specific value of θ that corresponds to the highest output of the likelihood function, and the correct interpretation of the MLE is that it is the value of θ that makes the observed data most likely [Fisher, 1922].

6.3.1 The ”Optimizing” Aspect

I have only identified one aspect of the MLE concept, *the ”optimizing” aspect*, which refers to the fact that the MLE is characterized as the method of finding the value that maximizes the likelihood function. This aspect leads to the definition of the maximum likelihood estimator in *definition 3*.

Definition 3. For a statistical model with sample space $\mathcal{X}$ and an observation $x \in \mathcal{X}$, one defines the maximum likelihood estimator as

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L_x(\theta)$$

given that this is well-defined [Lauritzen, 2023].

Notice that the MLE is not well-defined if the likelihood function doesn't have a maximum. However, it is always well-defined in the examples I will cover.

Technicalities: Since the logarithm is monotone and increasing I know that

$$\arg\max_{\theta \in \Theta} L_x(\theta) = \arg\max_{\theta \in \Theta} \ell_x(\theta).$$

When determining the MLE, I can maximize the log-likelihood function instead of the likelihood function.

6.3.2 The Grundvorstellung ”Solution to $\ell'(\theta) = 0$”

The ”optimizing” aspect, like for the likelihood function, gains its interpretive meaning from the Grundvorstellung ”relative measure.” This means that the MLE only makes sense in comparison with other hypotheses. Without this relative comparison, the ”optimizing” aspect can be misinterpreted and lead to an uncritical approach to ”the best estimate” even though the statistical evidence is low. The Grundvorstellung ”relative measure” is elaborated on in section 6.2.4, so the idea is not repeated here. Another interpretation of the ”optimizing” aspect focuses on the interpretation of the MLE as the solution to the equation $\ell'(\theta) = 0$.

In the case where the parameter space is continuous and the density functions are differentiable with respect to the parameter, I can find the maximum of the likelihood function using calculus [Watkins, 2011]. Let $S_x(\theta) = \frac{d}{d\theta}\ell_x(\theta)$ be the derivative of the log-likelihood function. In statistics this function is called the *score function*, and it has its own interests — however, I will only use it to maximize the log-likelihood function [Lauritzen, 2023].

The advantage of maximizing the log-likelihood function instead of the likelihood function is that it is much easier to find the derivative of the log-likelihood function. Returning to the example where I have n iid. observations, I know from equation 6.1 that the log-likelihood function is a sum of the log-densities, whereas the likelihood function is the product of the densities. Since it is easier to differentiate the sum than the product, the log-likelihood is preferred.

Hence to find the MLE, one sets the score function equal to zero and solve for θ. In statistics $S_x(\theta) = 0$ is called the score equation. Having found this extreme point, one should check that it is in fact a maximum either by inspecting the second derivative or by doing a function analysis [Watkins, 2011]. According to mathematics textbooks I have available in the Material List, upper secondary school students learn to find extreme points by solving the equation $f'(x) = 0$ and then, based on the monotonicity rule, construct a sign diagram for $f'(x)$ [Carstensen, 2021] [Brydensholt and Ebbesen, 2025a]. In this way, students can determine whether the extreme point is a maximum or a minimum.

Now, I will go through two examples where I find the MLE for different distributions, that may be relevant for upper secondary school students.

Example 2: MLE in the Binomial Distribution

I consider a sample x taken from a binomial distribution with unknown probability parameter $p \in [0, 1]$ and a known number of trials n. So x is an integer between 0 and n. The likelihood function is

$$L_x(p) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}.$$

Then I can calculate the log-likelihood by taking the logarithm of the likelihood

$$\begin{aligned}\ell_x(p) = \log(L_x(p)) &= \log\left(\binom{n}{x}\right) + \log(p^x) + \log((1-p)^{n-x}) \\ &= \log\left(\binom{n}{x}\right) + x \cdot \log(p) + (n-x) \cdot \log(1-p).\end{aligned}$$

The score function then becomes

$$S_x(p) = \frac{x}{p} - \frac{n-x}{1-p}$$

So the score equation is $\frac{x}{p} - \frac{n-x}{1-p} = 0$. Solving for p, I get

$$\frac{x}{p} - \frac{n-x}{1-p} = 0 \quad \Leftrightarrow \quad x(1-p) = (n-x)p \quad \Leftrightarrow \quad x = np \quad \Leftrightarrow \quad p = \frac{x}{n}.$$

Then the second derivative is

$$\frac{d^2}{dp^2}\ell_x(p) = -\frac{x}{p} - \frac{n-x}{(1-p)^2}.$$

Since this is negative for all values of p, I know that $p = \frac{x}{n}$ is the maximum likelihood estimate. I then write $\hat{p}_{MLE} = \frac{x}{n}$ [Ditlevsen and Sørensen, 2018].

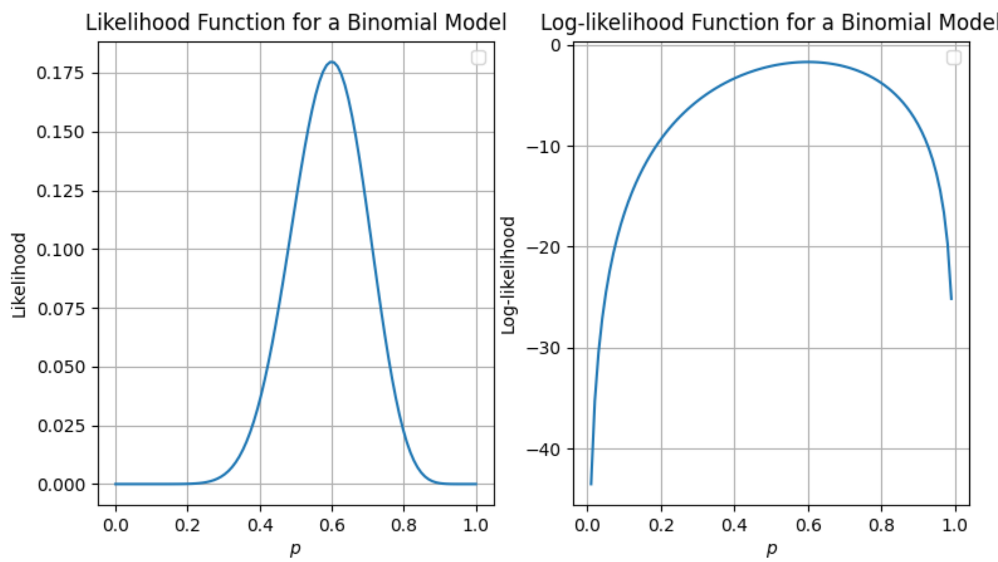


Figure 8: The figure shows the likelihood function and the log-likelihood function of the probability parameter p in a binomial model with $x = 12$ and $n = 20$ trials [Appendix B].

Example 3: MLE in the Normal Distribution

According to the syllabus, the normal distribution must be introduced to students with mathematics on A-level. This example may therefore be of interest, even though I have argued that the binomial distribution is generally the preferred for examples. It is in this example that I explicitly make use of the introduction to iid. random variables from section 5.1.4, which increases the level of complexity.

There are two parameters in the normal distribution, so either you could use multivariate calculus to maximize both parameters at the same time or you could assume one of the parameters is known in advance and then maximize the other parameter. Since multivariate calculus is not included in the A-level syllabus [Børne- og Undervisningsministeriet, 2024a], I will focus on the latter.

Let $x = (x_1, ..., x_n)$ be a sample where the observations are iid. normally distributed with unknown parameter $\mu \in \mathbb{R}$ and the known parameter $\sigma^2 \in \mathbb{R}_+$. The density function is given by

$$f_\mu(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$

for $y \in \mathbb{R}$. Then the likelihood function is

$$L_x(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right).$$

Now taking the logarithm, I get

$$\begin{aligned}\ell_x(\mu) = \log(L_x(\mu)) &= \sum_{i=1}^{n}\left(\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.\end{aligned}$$

So the score function is

$$S_x(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}n\left(\frac{\sum_{i=1}^{n} x_i}{n} - \mu\right).$$

I can then solve for μ in the score equation

$$\frac{1}{\sigma^2}n\left(\frac{\sum_{i=1}^{n} x_i}{b} - \mu\right) = 0 \quad \Leftrightarrow \quad \mu = \frac{\sum_{i=1}^{n} x_i}{n}.$$

The second derivative of the log-likelihood function is

$$\frac{d^2}{d\mu^2}\ell_x(\mu) = -\frac{n}{\sigma^2},$$

which is negative for all values of μ (since it is a negative constant) and hence the MLE is $\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$ [Ditlevsen and Sørensen, 2018].

The Grundvorstellung "solution to $\ell'(\theta) = 0$" can be transferred to the following competencies, that students can acquire.

- E1: Recognize that the MLE can be determined by solving $\ell'(\theta) = 0$.
- E2: Solve $\ell'(\theta) = 0$ in the case of a binomial model and a normal model. Either in the general setting of the binomial model or for a special case e.g. $n = 10$ and $x = 9$.
- E3: Apply tools from basic function analysis to argue that the solution to $\ell'(\theta) = 0$ is actually a global maximum.

6.4 Other Subject-Didactic Perspectives on MLE

I have now conducted a subject matter didactic analysis by identifying aspects and Grundvorstellungen related to the concepts: statistical model, likelihood function, and MLE.

In the following, the subject matter didactic analysis continues, and I will bring other perspectives to light, related to these concepts. The focus is on the subject-didactic contributions of likelihood functions and MLE, as legitimation for why this topic should be taught in upper secondary school. This is done through four perspectives: (1) relativity, (2) generality, (3) coherence and repetition, and (4) local connection.

6.4.1 The "Relativity" Perspective

The "relativity" perspective is the idea that likelihood functions itself carries information. While MLE can be misinterpreted and lead to an uncritical approach to 'the best estimate,' the likelihood function as "relative measure" can assist students statistical literacy. I have shown in relation to the Grundvorstellung "relative measure" how this approach can be obtained by visual inspection and calculation of the likelihood ratio. This can alternatively be done by likelihood intervals. I have chosen to exclude likelihood intervals from my analysis due to various considerations, such as use of the χ^2-distribution which is no longer a part of the upper secondary school syllabus [Børne- og Undervisningsministeriet, 2024a] and the prioritization of other perspectives. For those interested in likelihood intervals, Pawitan provides an introduction to likelihood-based intervals in his book *"In All Likelihood: Statistical Modelling and Inference Using Likelihood"* [Pawitan, 2001b].

Regarding the analysis in *'6.2.4 the Grundvorstellung "relative measure"'*, the likelihood function and MLE are legitimized by relativity because it:

- F1: Gives students a tool to compare two hypotheses relative to each other.
- F2: Provides students with tools to explore statistical evidence and the reliability of estimates. However, it does not offer definitive answers, but should be considered as a statistical analysis tool.

- F3: Takes the importance of the sample size n into account.
- F4: Has more than one mode of representation. Relativity is reflected both in the visual representation and in the calculation of the likelihood ratio.
- F5: Contributes overall to the development of students' statistical literacy: applying statistical knowledge and making statements about statistical evidence.

6.4.2 The "Generality" Perspective

The generality of MLE means that it can be applied in many different settings. This is due to the fact that MLE is a general technique that is not limited to a specific type of statistical model [Lauritzen, 2023]. Moreover, the method can be applied to both continuous and discrete data. This enables us to make exercises that are based on various different stochastic situations. There may be other advantages, and also some limitations associated with introducing more types of statistical models to upper secondary school students. I shall now show an example highlighting some limitations and advantages. I will later, in relation to developing a didactical design, discuss the limitations of teaching MLE in relation to the target group.

As already mentioned, the binomial and normal distributions are included in the A-level syllabus [Børne- og Undervisningsministeriet, 2024a]. Therefore, these distributions are obvious choices to work with in an optional topic about likelihood functions and MLE. Alternative distributions that can be explored include the exponential distribution, which can be used to model situations involving waiting times, the Poisson distribution, which can be used to model accidents or cases of illness over a period of time [Hadi and Sahib, 2023], and the hypergeometric distribution, which for example can be used to work with the Lincoln-Peterson method of mark and recapture [Pawitan, 2001a]. I will focus on the hypergeometric distribution since it highlights certain strengths and limitations.

Example 4: The Lincoln-Peterson method of mark and recapture

The Lincoln-Peterson method of mark and recapture is a technique used to count individuals in a population. The method starts by marking a subset of the population. Then a sample is taken, and the number of previously marked individuals in the sample is recorded. This method is commonly used to count wild animals. The proportion of marked and unmarked individuals provides an estimate of the population size N [Watkins, 2011].

The variables in the method: N is the total population size, t is the number of animals marked in the beginning of the experiment, k is the size of the sample (number of captured animals) and r is the number of marked animals in the sample. In the Lincoln-Peterson method t and k are chosen before the experiment and r is the observed number of marked animals. Then the purpose is to estimate the total population size N, which is unknown [Watkins, 2011]. The likelihood function for N is the hypergeometric distribution

$$L(N) = \frac{\binom{t}{r}\binom{N-t}{k-r}}{\binom{N}{k}}.$$

When determining the MLE, it is not a problem that data are discrete (as shown in the case with the binomial model). One can just use the Grundvorstellung "solution to $\ell'(\theta) = 0$" since the parameter space is continuous so the derivative $\ell'(\theta)$ makes sense. However, for the hypergeometric distribution where the parameter space is countable, one can no longer rely on taking the derivative with respect to N [Watkins, 2011]. Instead, one can use the ratio of the likelihood values for the successive value of N, namely $\frac{L(N)}{L(N-1)}$. To find the maximum of $L(N)$, one can determine when this fraction is strictly smaller than 1 and when it is strictly larger than 1. By the calculations in appendix D, one can see that

$$\frac{L(N)}{L(N-1)} > 1 \quad \Leftrightarrow \quad N < \frac{tk}{r}.$$

This shows that the function $L(N)$ is maximized at $N = \lfloor \frac{tk}{r} \rfloor$.
Then the MLE is $\hat{N}_{MLE} = \lfloor \frac{tk}{r} \rfloor$ [Watkins, 2011].

The point of introducing this method is foremost to work with an engaging example with data grounded in the real world. The Lincoln-Peterson method is an example of how one can contextualize the likelihood approach. As shown in appendix D, the Lincoln-Peterson method is derived using mathematical concepts that occur in the school textbooks I have looked at in MATERIAL LIST. These concepts include the factorial, arithmetic rules of fractions and inequalities. Nonetheless, I think the calculations involved in appendix D are quite heavy and the conclusion that $L(N)$ is maximized at $N = \lfloor \frac{tk}{r} \rfloor$ is conceptually difficult. I think that the method poses some limitations in terms of calculation and conceptual difficulty. This leads me to the next key point: Not all statistical models are equally suitable for applying the MLE when teaching upper secondary school students. If one still wishes to work with such examples (e.g., the Lincoln-Peterson method), I suggest the maximum likelihood estimator can be identified graphically instead of though calculations. Alternatively, it can be used in student projects where the subject can be explored in more depth.

The advantages of the "generality" perspective can be summarized as follows:

- G1: MLE is a technique that allows for working with various types of stochastic situations and statistical tasks.
- G2: The problems can be motivated by real-world examples.

The analysis has shown that the generality of MLE has some limitations in an upper secondary school context, since some statistical models are difficult to work with.

6.4.3 The "Coherence and Repetition" Perspective

This perspective is about the ability to create strong connections between the syllabus in mathematics and statistics. As mentioned in the section about the theoretical framework, statistical reasoning and mathematical reasoning has its own distinct characteristics. Despite these differences, both subjects belong to the same upper secondary school subject and syllabus. The positioning of statistics within upper secondary school mathematics sets certain boundaries for the statistical content. For instance, it seems essential to create a strong link between mathematics and statistics [Scheaffer, 2006]. The situatedness of statistics in school mathematics as is the case in Danish upper secondary school is illustrated in Figure 9.

A strong coherence between mathematics and statistics means that the two fields are taught in a way where they strengthen each other and the "marriage" between them is legitimized [Scheaffer, 2006]. For instance, students should understand why statistics is a part of the mathematics syllabus so that it does not feel like a disconnected topic unrelated to the rest of the core content in the syllabus, but they should also know the distinction between mathematics and statistics [Scheaffer, 2006]. This means that students both perceive the demarcation between statistical reasoning and mathematical reasoning and perceive a general coherence between statistics and mathematics, allowing students to identify shared traits and fundamental ideas common to both disciplines [Scheaffer, 2006]. This can according to education researcher Randall E. Groth be achieved by that a shared problem space between the disciplines become a larger part of the content in school mathematics [Groth, 2015].

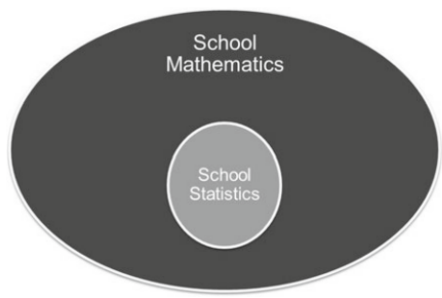


Figure 9: The Situatedness of Statistics in School [Weiland, 2019].

In the following, I will argue that MLE creates a strong connection between the two branches of upper secondary school mathematics: functions and calculus (classes of functions, function analysis and differential calculus) and statistics and probability. In doing so, I also show that MLE creates coherence between statistics and other fields of mathematics.

The method of MLE touches fundamental ideas that are important to both mathematics and statistics such as modelling, representation and estimation. The examination of fundamental ideas underlying MLE builds on my own analysis, but some more general notions of modelling, representation and estimation relies on Gail Burrill and Rolf Biehlers article *"Fundamental Statistical Ideas in the School Curriculum and in Training Teachers"* [Burrill and Biehler, 2011].

- **The fundamental idea of modelling.** Within mathematics, the notion of a model is a mathematical object (a function), interpreted as a simplification or idealization of reality [Burrill and Biehler, 2011]. This idea is also central to statistics and the "statistical model" concept. A difference is that mathematical modelling tends to ignore variability and uncertainty. This, on the other hand, is one of the key ingredients of the fundamental idea ”modelling” in statistics [Burrill and Biehler, 2011]. The likelihood approach underscores the fundamental idea of modelling in both a mathematical and a statistical sense. Mathematical modelling gives the mathematical language to formalize the structural parts of the statistical model. Modelling in a statistical sense deals with the stochastic part of the model and its contextualization.

- **The fundamental idea of representation.** The fundamental idea of representation in mathematics is about shifting between different modes of representation [Burrill and Biehler, 2011]. In relation to likelihood functions, this includes the semantic understanding of the function as a ”reversed density” versus the visual representation of the likelihood function. The fundamental idea of representation in statistics is as a key tool for analysis [Burrill and Biehler, 2011]. In relation to likelihood functions, I use visual inspection of the likelihood graph to analyse the grundvorstellung ”relative measure”.

- **The fundamental idea of estimation.** The fundamental idea of estimation in the mathematical sense is done by proportional reasoning: this means to transmit the features of the sample directly to the whole population. In other words, one assumes a perfect proportional relationship between the two [Burrill and Biehler, 2011]. Calculating the MLE and choosing $P_{\theta_{MLE}}$ as the underlying data-generating function behind some phenomena support proportional reasoning. The fundamental idea of estimation in statistics puts emphasis on data and the uncertainty associated with it. Some of the key ingredients in this are: (1) The intuition of ”weakness”, that the information solely comes from data and therefore it might not be possible to say anything certain about the population. (2) The stochastic component in the statistical model: acknowledging that a sample is non-reproducible, meaning there will always be variations from sample to sample, and (3) The inclusion of relative likelihood to assess the reliability of an estimate.

I conclude that the likelihood function and MLE relies on the fundamental ideas: modelling, representation and estimation — all essential to statistical reasoning and mathematical reasoning. In addition to connections created by fundamental ideas, MLE also establishes more concrete connections to other topics or concepts in the official syllabus. I will used these connections to justify why it is within reach to teach MLE and how repetition plays an important role in the perspective ”coherence and repetition”. The topic is recognizable due to its many links to the core mathematical (and statistical) content in syllabus, which may give students a sense of familiarity and confidence in the subject. This I will explain through a connection diagram in figure 10, which I have developed using the A-level syllabus, Material List and my previous analysis on the ”MLE” concept.

The column in the middle of Figure 10 consists of four central steps in determining the MLE building on the Grundvorstellung ”solving $\ell'(\theta) = 0$”:

A: Write down the likelihood function.

B: Take the natural logarithm of the likelihood function.

C: Obtain the score function and solve the score equation.

D: Check that the MLE estimator corresponds to a maximum.

The left and right columns show various connections, illustrating the core content of the A-level syllabus occurring in the central steps. These include core content such as function analysis, differential calculus, probability, and logarithmic functions [Børne- og Undervisningsministeriet, 2024a]. Thus, the diagram demonstrates why MLE could be considered as a shared problem space between mathematics and statistics

- with (A) writing down the likelihood function (as well as interpretations on likelihoods and MLE based on the Grundvorstellung ”relative measure”) relying on the probabilistic and statistical knowledge of sampling, statistical models and stochastic.
- and with (B), (C) and (D) relying on calculus and function analysis to solve a optimization problem.

Therefore, MLE is a statistical topic building bridges to the mathematical core content in the syllabus. The connection diagram in Figure 10 especially shows how the branches of statistics and functions are being connected though MLE. The year in parenthesis is the year the concept is typically taught in upper secondary school. This information is based on the Material List.

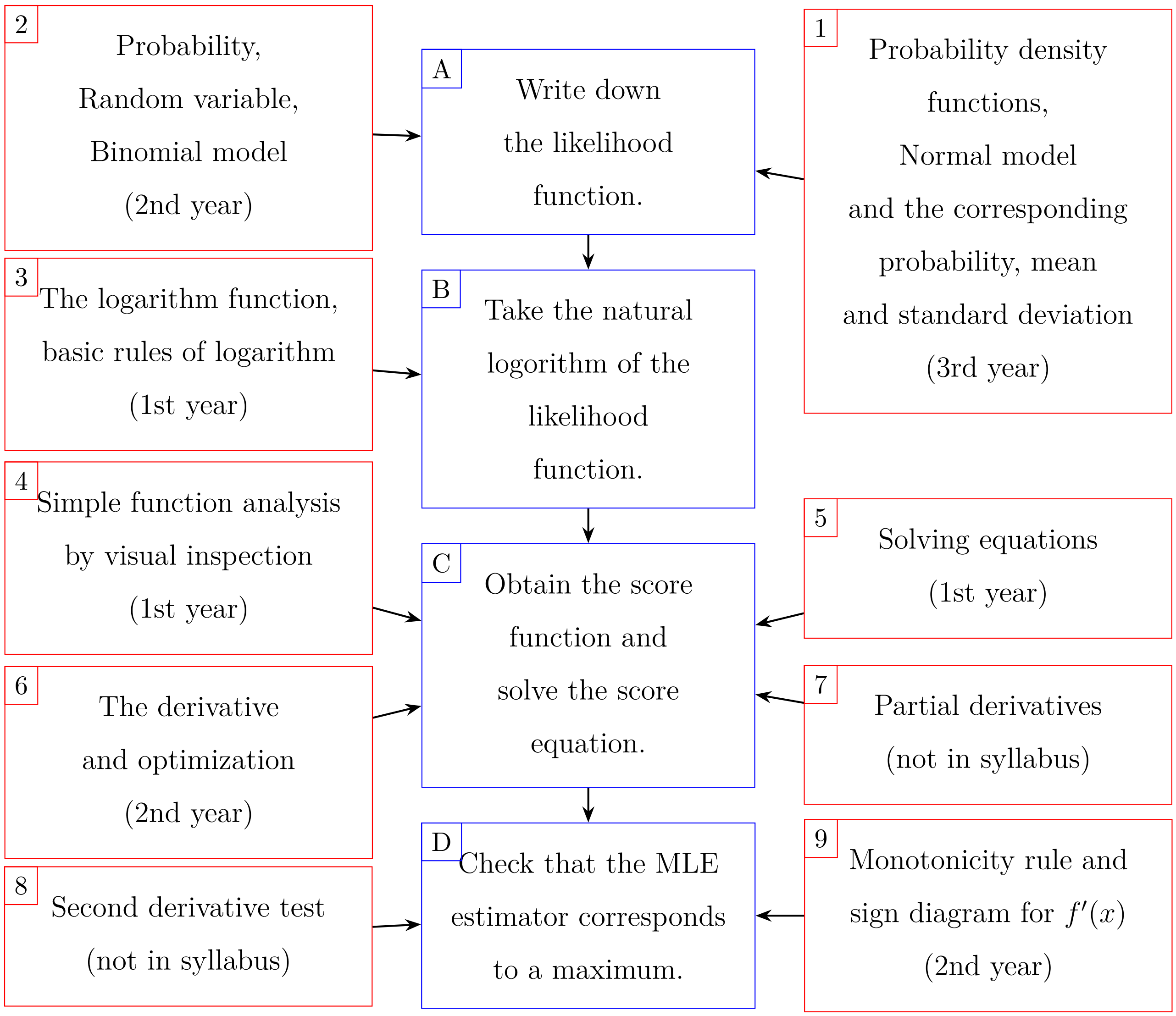


Figure 10: Connection Diagram of teaching MLE in upper secondary school [Appendix B].

The "coherence and repetition" perspective legitimize likelihood functions and MLE as "knowledge to be taught", since it

- H1: Supports fundamental ideas that are prevalent in both statistics and mathematics.
- H2: Creates strong connections between different branches within upper secondary school mathematics.
- H3: Repeats and extends some topics and concepts from the official syllabus. The recognition may give students a sense of familiarity and confidence.

6.4.4 The ”Local Connection” Perspective

In the ”local connection” perspective, I will continue looking into the subject-didactic structure of MLE. I will argue that different statistical topics such as linear regression, descriptive statistics, probability and statistical inference are often taught as more or less separate topics, and depending on the teaching approach, students may struggle to see the interconnections. Furthermore, I show that MLE can create local connections. For example, one can show that the least squares method is the same as calculating the maximum likelihood estimate for ordinary linear models.

The structure of almost every introductory statistics course is to first start with descriptive and exploratory data analysis, then move into probability, and finally go to statistical inference [Biehler, 1994]. Biehler warns us that the danger of a syllabus with such a structured progression is that students get the impression that *“EDA (Exploratory Data Analysis), probability and inference statistics seem to be concerned with very different kinds of application with no overlap”* [Biehler, 1994, p. 16].

The connection between probability and statistical inference in the likelihood approach is quite simple. The likelihood function is derived from a probability density function and can be used as a tool to gain statistical inference. Likewise, a solid foundation of basic probability concepts is necessary to introduce the ”statistical model” and ”likelihood function” concept as seen in section 6.1 and 6.2.

One can show that MLE also can create local connection to ordinary linear regression, which is contained in the syllabus as core content [Børne- og Undervisningsministeriet, 2024a]. Linear regression is used to model the statistical association between two variables x and y [Watkins, 2011]. The linear regression model is

$$y_i = ax_i + b + \varepsilon_i,$$

where one considers n observations with $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$. The ε_i are called the residuals and are assumed to be iid. normal-distributed with mean zero.

I will now show that the MLE is equivalent to the least square method for ordinary linear regression. Since $\varepsilon_i = y_i - (ax_i + b)$, one consider it a random variable with parameters a, b and σ^2. The parameter σ^2 is considered fixed, so the joint density for $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is

$$f_{a,b}(x, y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right).$$

The likelihood function for ε then becomes

$$L_{x,y}(a, b) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right).$$

Taking the logarithm of the likelihood function

$$\begin{aligned} \ell_{x,y}(a, b) &= \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (ax_i + b))^2}{2\sigma^2}\right)\right) \\ &= -n \cdot \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (ax_i + b))^2 \end{aligned}$$

The first term is constant with respect to a and b and since each term in the sum is positive, the log-likelihood function is maximized when the sum is minimized. So, I conclude that finding the MLE is the same as minimizing the sum of squares $\sum_{i=1}^{n} \varepsilon^2$ [Watkins, 2011].

The advantages of the ”local connection” perspective can be summarized as follows:

- I1: MLE connects the probability topic and the statistical inference topic.
- I2: MLE can be considered as a generalization of the ordinary least square method. Therefore, MLE creates a local connection between likelihood functions and the linear regression topic.

In order to refer to the conclusions in later section, the summaries above have been given a reference label $A1, A2, A3, B1, ..., I1, I2$.

6.5 Overview: Subject Matter Didactic Analysis

In figure 11, I provide an overview of my results in the subject matter didactic analysis. The figure is a flowchart intended to help organize the analysis. Among other things it illustrates the relationships between the various concepts, aspects, and Grundvorstellungen.

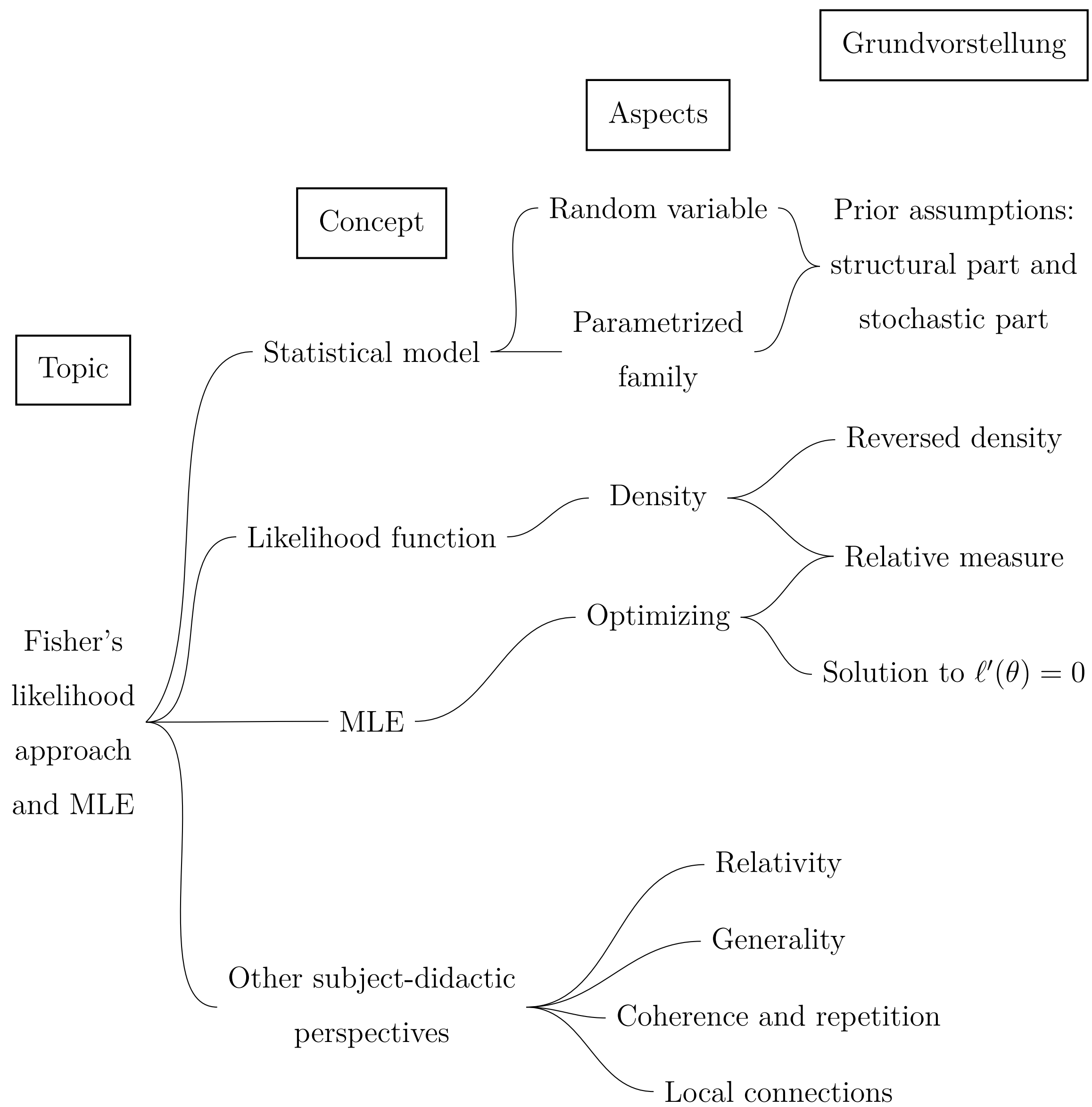


Figure 11: Overview of the Subject Matter Didactic Analysis [Appendix B].

7 Didactical Design

The didactical design consist of development and analysis of a mathematics textbook chapter on likelihood functions and MLE. The textbook chapter can be found in *Appendix A* in Danish and the textbook chapter can with advantage be accessed before reading any further. This is due to that the textbook chapter was developed in parallel with the following analysis on the didactical design. Therefore, I explicitly refer to the textbook chapter explaining my considerations and design choices along the way. I have approached the didactic transposition of a textbook chapter through the methodology of didactical engineering. The analysis is divided into two parts. The first part focuses on the implementation of the design by doing an institutional analysis *(section 7.1)*. The second part focuses on the praxeological organisation and the task design by doing an a priori analysis *(section 7.2)* [González-Martín et al., 2014]. As mentioned in section 2, this thesis does not use the complete methodology of didactical engineering.

The didactical design should not be seen solely as the textbook chapter, which Kang and Kilpatrick describe as a static form of knowledge. The design also involves the application of the chapter into a dynamic form of a teaching practice and an environment [Kang and Kilpatrick, 1992]. Likewise, in the following my considerations and design choices both concern the static part of the product (the textbook chapter) and the dynamic part of implementing the textbook chapter (a contextualization). Since the design has not been tested in upper secondary school, it can only give rise to a pseudo-contextualization. This implies that the analysis engages with hypothetical (or imaginary) students, trying to forecast teaching activities in a real environment [Kang and Kilpatrick, 1992].

The institutional and a priori analyses are carried out in terms of mathematical praxeologies (ATD) and statistics education research. Besides these theories, the analysis of the didactical design builds on the conclusions from the subject matter didactic analysis (section 6).

7.1 Part One: Institutional Analysis

The following analysis deals with the implementation of the design and explores the opportunities and limitations of the implementation in upper secondary school.

This is done by considering the target group of the design, students' prerequisites and the institutional conditions such as the national syllabus and textbooks, and by proposing a design implementation, as an optional topic.

7.1.1 The National Syllabus in Mathematics

The national syllabus in mathematics (stx), developed by the Danish Ministry of Children and Education, outlines the official requirements and recommendations for the knowledge to be taught in upper secondary school [Børne- og Undervisningsministeriet, 2024a]. As late as in 2024, a new mathematics syllabus was introduced, which is currently in force.

The national mathematics syllabus consist of three distinct course instructions, one for each level of school mathematics: A, B, and C. In my investigation, I have used the new (2024) syllabus for A-level, since A-level mathematics students is the target group for my didactical design. Any syllabus in Danish upper secondary school consists of four sections, first, the identity and purpose of the subject, second, the core content, third, the organization of teaching, and finally, evaluation. In addition to the syllabus, a supplementary guide (Vejledning til læreplan) is provided, which offers much greater detail, particularly regarding the content of the subject [Børne- og Undervisningsministeriet, 2024b].

7.1.2 The Material List

Textbooks in mathematics also play an important role in shaping conditions and constrains surrounding the knowledge taught in school mathematics. This is because, a course in mathematics often follows a certain textbook, and the choice of book heavily influence the teaching practices [Kang and Kilpatrick, 1992].

The selected textbooks in the Material List in section 6.1 represents different questions regarding conditions, which are useful in my investigation. Some books are chosen to distinguish what core content is ”typically” taught in respectively 1st, 2nd, and 3rd year of upper secondary school: ”Mat A2 stx” [Carstensen, 2021] and ”Lærebog i matematik A3” [Brydensholt and Ebbesen, 2025b]. Some books are chosen according to the focus on statistics: ”Matemat10k : statistik” [Agermose Jensen and Timm, 2014], ”Sandsynlighedsregning og statistik” [Clausen et al., 1997] and ”Statistik C” [Gråbæk et al., 2025]. Other materials represent books with a higher level of complexity or a more theoretical approach: ”Hvad er matematik?” [Grøn, 2015] and ”Højniveaumatematik. Bind 2” [Hebsgaard and Sloth, 1999]. Having access to different mathematics textbooks has also been useful as inspiration for writing a textbook chapter myself.

7.1.3 The Target Group

I have determined that the didactical design is intended for students at upper secondary school (stx) with mathematics at a high level (A-level). Furthermore, it is preferred that the students are on there third and final year (3.g) of Danish upper secondary school. The decisions regarding the target group has been made due to the theoretical nature of the topic, which in my own experience is first taught at university. I therefore assume that successfully communicating likelihood functions and MLE to upper secondary students will require students to have as much prior knowledge as possible. The subject matter didactic analysis support my assumption regarding the level and complexity of the topic. In particular, the perspective ”coherence and repetition” show that MLE is an add-on to a lot of the core content in the mathematics national syllabus.

Referring to the connection diagram in Figure 10 (section 6.4), the calculus and function analysis required to perform the method of MLE are often taught in the second year of upper secondary school (according to the Material List). However, density functions and the normal distribution are usually not introduced until the third year [Brydensholt and Ebbesen, 2025b]. As the subject matter didactic analysis show, density functions are important for characterizing and understanding the ”likelihood” concept by the ”density” aspect and the Grundvorstellung ”reversed density”. The design is therefore not intended for first or second year students, since they do not have all the prerequisites.

7.1.4 The Basics of Statistics

In this thesis, MLE is not proposed as an alternative to the current syllabus in statistics and probability, but a supplementary add-on. I think it is important that students, before approaching a theoretical statistical topic like MLE, already possess an understanding of basic statistical concepts like *frequency, sample, population, central tendency, distribution, randomness and representativeness.* Likewise, statistics education researchers George W. Cobb and David S. Moore in their paper *"Mathematics, Statistics, and Teaching"* [Cobb and Moore, 1997] argue that an overly theoretical and formal way of doing statistics is not a instructive way of approaching the basic concepts. In general, less formal and software-driven methods for teaching statistical inference are commonly emphasized in statistics education research instead of formal approaches, especially when targeting students younger than 16 [Makar and Rubin, 2017] [Batanero and Borovcnik, 2016]. Since the target group for this design consists of students aged 17–19, a formal approach may be meaningful if the students have the right prerequisites. The basic statistical concepts should therefore be prerequisites, and not a learning outcome of the didactical design. If the students lack the proper statistical prerequisites, they may not be able to understand what is being taught about MLE.

It is also necessary that the students have a teacher who puts emphasis on statistics and the basic statistical concepts mentioned above [Scheaffer, 2006]. Both Linda Gattuso and Maria Gabriella Ottaviani in their paper *"Complementing Mathematical Thinking and Statistical Thinking in School Mathematics"* [Ottaviani, 2011] and Richard L. Scheaffer in his paper *"Statistics and Mathematics: On Making a Happy Marriage"* [Scheaffer, 2006] address how mathematical reasoning and fundamental mathematical ideas dominate school mathematics. Statistical ideas and ways of thinking are less prominent and may be neglected. I do not believe the suggested design can be implemented in an environment where statistics have been neglected or de-emphasized, due to the complexity of the subject matter.

7.1.5 Students Prerequisites

Students should have been through courses about (1) descriptive statistics, (2) probability and combinatorics, (3) the binomial distribution and binomial test, and (4) the normal distribution and density function. According to the national syllabus, these courses should ideally provide the following statistical prerequisites.

- Empirical materials and statistical descriptors such as the mean and the median.
- The concepts: sample, population and representativeness.
- Stochastic variables and the notation $X \sim bin(n, p)$.
- Events and independent events and have encountered the latter in relation to tasks using multiplication of probabilities (known as the multiplication principle).
- The probability mass function of the binomial distribution.
- Estimation of the probability parameter p in a binomial model as the ratio between the number of successes x and the number of trials n.
- Statistical hypotheses in connection with binomial tests and the notation $H_0 : p = p_0$.
- The probability density function of the normal distribution.
  [Børne- og Undervisningsministeriet, 2024a]
  [Børne- og Undervisningsministeriet, 2024b]

Beyond the statistical prerequisites outlined above, students are also expected to have prior knowledge in the areas of calculus and function analysis. In order to understand the Grundvorstellung 'solving $\ell'(\theta) = 0$', students must be able to determine the log-likelihood function using logarithmic rules, differentiate the function, solve the score equation, and verify that the solution indeed is a global maximum [E1-E3]. I have identified the following prerequisite knowledge in the syllabus that students are expected to have.

- Basic rules of logarithmic operations.
- Differential calculus, including arithmetic rules for derivatives.
- Graphical interpretation of monotonicity and extrema, as well as the use of f' to determine extrema and monotonicity, and to construct signs diagrams for f'.
  [Børne- og Undervisningsministeriet, 2024a]

In addition to this list, the supplementary topic in the second year (2.g) of upper secondary school could advantageously involve summation (the notation $\sum$). If students have encountered summation notation before, then it will be easier to introduce the product notation ($\prod$). However, if students have not previously encountered summation notation, this may be a limitation of the didactical design.

7.1.6 MLE as Optional Topic

I suggest implementing likelihood functions and MLE in upper secondary school as an optional topic. This is possible since the new syllabus (2024) allocates 10 percent of the total teaching time to supplementary content [Børne- og Undervisningsministeriet, 2024a]. According to the syllabus, the following must apply to the supplementary content.

> *"Det supplerende stof, der skal udfylde mindst 10 pct. af undervisningstiden, skal uddybe arbejdet med kernestoffet, indeholde nye emner eller metoder og perspektivere faget med vægt på faglig argumentation."*
> [Børne- og Undervisningsministeriet, 2024a, p. 2]

The supplementary content must meet certain requirements, e.g. it should introduce some new content, while at the same time expand, strengthen or elaborate on the core content in the national syllabus [Børne- og Undervisningsministeriet, 2024b].

My investigation has shown that likelihood functions and MLE strengthen the core content in statistics, as well as content across different branches of school mathematics by the "coherence and repetition" perspective framed in H1-H3. Several core concepts from the national syllabus are extended in the likelihood topic. Among these core concepts are density functions and statistical models. In the national syllabus, the concept of density function is described as integrals used to determine "interval probabilities" [Børne- og Undervisningsministeriet, 2024b]. With C1 and C2, the concept of the density function is expanded by characterizing and giving meaning to the likelihood function. By the "statistical model" concept, framed in A1–A3, the concept of a statistical model is extended from the concrete binomial model to a general statistical model. This generalization is stated in the "generality" perspective, framed in G1 and G2, enabling statistical investigations to be applied to a wide range of models. The likelihood approach can thus both elaborate on the work with the binomial and normal distributions and be extended to other distributions as well. I choose that the topic will be positioned after teaching the normal distribution and its density function (in 3.g), so that it becomes a natural extension of this.

In general, the likelihood approach can be seen as an extension of statistical inference and hypothesis testing, which in the national syllabus consist of the binomial test by normal approximation [Børne- og Undervisningsministeriet, 2024b]. By the ”relativity” perspective, framed in F1–F5, students can broaden their toolkit for statistical analysis and develop a deeper understanding of how to interpret statistical evidence.

The national syllabus also states that the supplementary content should put the subject of mathematics into perspective, with an emphasis on mathematical reasoning [Børne- og Undervisningsministeriet, 2024a]. In this context, likelihood functions and MLE illustrate how mathematics (including calculus and classical algebra) can be applied in statistical investigations and thereby have to do with real-world phenomena. The content of MLE can therefore shed light on the interplay between mathematical and statistical reasoning. This is because the derivation of the maximum likelihood estimate relies on mathematical reasoning, through the deductive derivation of the log-likelihood function and the identification of its global maximum. However, it is based on data and statistical assumptions — both elements that belong to statistical reasoning and involve an inductive approach. Similarly, the analysis and interpretation of the likelihood function and the maximum likelihood estimate, based on the Grundvorstellung of “relative measure,” as framed in D1–D3, cannot be understood solely through mathematical reasoning. Rather, my investigation in the subject matter didactic analysis shows that the interpretation of MLE should involve critical statistical reasoning and statistical literacy (F1-F5).

This perspective, relating types of reasoning, can more generally be understood as the interplay between two dimensions of school mathematics: pure mathematics (with mathematics as a structure) and the practice-oriented dimension (with mathematics as a model), as described by Jens Christian Larsen and Kasper Bjering Søby Jensen in their paper ”Metoder og videnskabsteori i, med og om matematik” [Larsen and Jensen, 2019]. Within the likelihood approach, the practice-oriented dimension is represented through statistical reasoning, where the real-world contextualization is both natural and essential in the way data is used. Mathematics as structure is used for deriving the maximum likelihood estimate. A supplementary topic about MLE can provide an interplay between the deductive and cumulative structure of mathematics and the practice-oriented and inductive nature of statistics. Hence MLE is a shared problem space [Groth, 2015].

Since the topic of MLE both elaborates on the core content and adds perspectives on mathematics with emphasis on mathematical reasoning, it falls within the requirements for supplementary content in the national syllabus. It is worth noting that the official guidelines specify that such perspectives may be internal to mathematics (and statistics), meaning there is no requirement for the supplementary content to be seen in perspective to other upper secondary school subjects [Børne- og Undervisningsministeriet, 2024b].

Finally, I consider the practical implementation of such optional topic in a teaching course. The percentage allocated to supplementary content is relatively high. Assuming a total of 375 hours over three years, and that each lesson lasts 1 hour and 30 minutes [Børne- og Undervisningsministeriet, 2024d], this corresponds to the possibility of offering optional topics of around 8 lessons per year. Some of this time will probably be used for feedback on assignments and tests. Based on the textbook material I have written [Appendix A], I propose a teaching sequence consisting of five lessons. How the chapter and designed activities should be structured across the five lessons are explained in part two of this analysis.

7.1.7 Limitations

In communicating an academic and theoretical topic (MLE) to upper secondary students a number of limitations arise. The limitations in this section concerns the fact that both teachers and students may not have the prerequisites or necessary tools regarding MLE.

According to A. Harradine, C. Batanero and A. Rossman in "Students and Teachers' Knowledge of Sampling and Inference" [Harradine et al., 2011], there have been very few studies on mathematics teachers' knowledge of statistics. In their paper, they conclude that teachers often face many of the same difficulties in understanding statistics as their students do.

Some of the difficulties mentioned in [Harradine et al., 2011] relate to the following.

— Not understanding the difference between randomness and representativeness.

— Believing that statistical evidence can be obtained based on very small datasets.

— Failing to distinguish between the population distribution and the sample distribution.

— Confusing the rejection of a hypothesis with the hypothesis being definitively wrong.

— Struggling to interpret both p-values and confidence intervals correctly.

Even though this analysis does not apply directly to MLE, it does support the fact that mathematics teachers at this level encounter some challenges in teaching probability and statistics [Harradine et al., 2011]. The difficulties that teachers experience in a classic statistical approach (with the binomial test and confidence intervals) would likely also apply to the likelihood approach. In fact, the difficulties may be even bigger, given that the topic of MLE is unfamiliar in the context of upper secondary education.

> *"Uncertainty and statistical inference are challenging ideas for teachers, just as they are for the general population."* [Batanero and Borovcnik, 2016, p. 342]

A possible limitation would be to find a teacher who is both willing and qualified to teach MLE. If the teacher is unfamiliar with the topic, or lacks the necessary statistical qualifications, it could demand considerably more preparation than usual. One can also imagine that some teachers are simply less interested in statistics, and unlike myself (who believes statistics is highly important) may feel that the syllabus already includes more than enough statistics. These limitations strongly depend on teachers' attitudes toward teaching statistics and teachers' training in statistics [Batanero and Borovcnik, 2016].

According to the current study programs at Danish universities, upper secondary school teachers cover MLE in a mandatory university course [Københavns Universitet, 2025] [Aarhus Universitet, 2025] [Det Naturvidenskabelige Studienævn SDU, 2025]. Since I have not examined teachers' knowledge of likelihood functions and MLE, it is uncertain how much teachers actually remember about this topic and whether they consider it difficult.

The following limitations concerns the students' prerequisites. In the subject matter didactic analysis, I found that many of the key concepts related to likelihood functions and MLE are not included in the national syllabus or in the available textbooks form the MATERIALS LIST. This created several limitations and challenges, which I have had to navigate while writing the textbook chapter.

As I concluded in the subject matter didactic analysis, one limitation is that students may not know what a statistical model is. As mentioned in section 6.1, the "statistical model" concept is central to introducing the "likelihood function" concept. Therefore, I have decided that statistical models need to be addressed explicitly in the textbook chapter. This is done using the "random variable" aspect to characterize statistical models. This aspect is chosen over the "parametrized family" perspective partly because some textbooks already use the random variable framework [Clausen et al., 1997], and partly because it uses the notation $X \sim bin(n, p)$, which is included in the core content in the syllabus [Børne- og Undervisningsministeriet, 2024b].

Another limitation, found in my investigation, is that students do not know about iid. random variables. In the textbook chapter I do not explain iid. random variables, but I state the fact that the joint density is a product of the marginal densities. This gives rise to a blind alley, but the student have seen the multiplication principle which can be used as an explanation.

A third limitation is that the parameter space might not be continuous. Models with a countable parameter space such as the hypergeometric distribution makes it impossible to determine the maximum likelihood estimate using standard differential calculus. This creates limitations regarding which models are practical to work with. To work around this limitation, the students are only asked to find the MLE of the hypergeometric distribution through graphical means. However, I chose to primarily focus on the normal and the binomial distribution in the chapter, so this limitation is not crucial.

Finally, it is a limitation that students do not know multivariate functions. In my analysis, I found that this becomes relevant, e.g. if one wishes to estimate both the mean and the standard deviation in a normal model using the MLE method. Therefore, I decided that in the chapter's treatment of the normal distribution, the standard deviation would be treated as fixed, and only the mean would be considered an unknown parameter. This allowed me to avoid working with functions of multiple variables.

Some of these limitations could potentially be addressed in a written student project (SRP) on the likelihood function. In such a project, students are expected to engage more deeply with new knowledge, e.g. learning about iid. random variables or multivariate functions [Børne- og Undervisningsministeriet, 2024c].

7.2 Part Two: A Priori Analysis

In the following, I conduct an a priori analysis as part of the didactical design, where I explain the didactical engineering behind the textbook chapter and the corresponding design activities in more details. The a priori analysis will include considerations regarding the structure of the lessons and the task design. The structure of the design concerns fundamental design choices regarding the selected content and how the content is approached through tasks that students can solve. The investigation of task design is about students' possible strategies and difficulties in solving mathematical tasks. I am analyzing some selected tasks from the textbook chapter and considering hypothetical solutions.

I make a number of decisions regarding how to organize the design. One consideration, described in the previous section (7.1), is concerned with the scope of the design. I decide that the textbook chapter should correspond to an optional topic and consist of five lessons. Also, I decide that the chapter should communicate likelihood functions and MLE to students by teaching them to derive the maximum likelihood estimate and to compare two hypothesis by their likelihoods. I have chosen, that the textbook chapter should build on the binomial distribution and the normal distribution, recalling students prior knowledge about the distributions. I choose to place the optional topic about MLE as a continuation of the teaching sequence about the normal distribution. Therefore, the normal distribution is included as the first example in the textbook chapter, serving as a natural extension of the statistical core content on third year. The maximum likelihood estimate for the normal distribution is derived in the textbook chapter, while the students derive the maximum likelihood estimate for the binomial distribution themselves as an exercise. Finally, the first half of the design is about MLE, and the second half is about comparing hypothesis relying on my investigation of the Grundvorstellung "relative measure". This means that students should acquire the competencies (D1) comparing two hypothesis graphically and by calculation, (D2) determine the likelihood ratio and (D3) be familiar with effect of the sample size n on the likelihood curve.

To structure my didactical design, I have chosen to build a praxeological reference model using the tools from ATD. The praxeological reference model has been chosen as an organizational tool for my design because it takes into account the epistemological considerations, which are essential when transforming "scholarly knowledge" into a teaching design with "knowledge to be used" by students [Kang and Kilpatrick, 1992]. In addition, the framework for the reference model is partly based on my own result from the subject matter didactic analysis, in which I have outlined the organization of the scholarly knowledge and found what competencies students can acquire. Moreover, the praxeological reference model developed by R. Hakamata, K. Otaki, H. Fukuda, and H. Otani in the article *"Statistical modelling in the Brousseaunian guessing game: A case of teacher education in Japan"* [Hakamata et al., 2022] has inspired the structure of my own model.

In the article about the Brousseaunian guessing game [Hakamata et al., 2022], their reference model consists of an inferential statistical praxeology $\mathscr{P}_{IS}$, which gives rise to sub-praxeologies that specify the different schools within statistical inference, e.g. "Frequentist school", "Fisherian school" and "Bayesian school". Since my thesis focuses solely on the Fisherian approach to inferential statistics, the inferential statistical praxeology is just the Fisherian praxeology $\mathscr{P}_{IS} = \mathscr{P}_{Fisher}$. Alongside the inferential statistical praxeology, the reference model in the article [Hakamata et al., 2022] includes a practice-oriented component, consisting of the inquiry of a guessing game denoted as the experimental praxeology. The practice-oriented part of my design is very different, as it does not encourage students' autonomous inquiry of a game, but rather follows a more traditional textbook format with exercises. The practice-oriented part in the didactical design consists of the different contextualizations in the exercises and examples included in the chapter. I will denote it by

$\mathscr{P}_{Context}$. Hence, my reference model consists of two sub-praxeology with the structure: $\mathscr{P}_{ref} = \mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context}$.

Before I describe the five lessons, I will elaborate on the praxeology $\mathscr{P}_{Context}$ by giving the context of the exercises and examples. In my design, I have aimed to work with different forms of contextualization to motivate the content and to use the "generality" perspective. This is reflected in the following examples and exercises used in the textbook chapter.

(1) An example on Danish women's foot lengths and the normal distribution. This example is inspired by Bo Markussen's material on the normal distribution [Markussen, 2020]. I myself added the likelihood approach to the context given in Markussen's material.
(2) An example about opinion polls for the Danish political party, the Social Democrats. This is a popular example in school textbooks, when introducing the binomial test. It is used in [Grøn, 2015], [Clausen et al., 1997] and [Carstensen, 2021]. I myself added the likelihood approach to the context of opinion polls and votes for the Social Democratic Party.
(3) An exercise on the Lincoln-Peterson method of mark and recapture (this exercise is about ice birds), which is inspired by an example in [Pawitan, 2001b] about badgers.
(4) Classical examples and exercises with coin tosses and dice rolls. These examples and exercises was primary inspired by an example by [Etz, 2018].
(5) An exercise about a sauna and a man called Earl, which is also inspired by [Etz, 2018].

In the a priori analysis, I will often refer to the general formulation of a task or technique in purely theoretical terms of $\mathscr{P}_{Fisher}$, but the version students encounter in the textbook chapter will typically be presented within a contextualized setting of $\mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context}$, following Kirsch's paradigm of 'making accessible by including the surroundings of mathematics'.

In the following, I will conduct an a priori analysis the five lessons. Exercises in the textbook chapter can in principle be done without digital tools. However, for some exercises it is recommended that students use their CAS-tool as a calculator, but not anything more advanced.

7.2.1 Lesson 1: Introduction

Refer to appendix A '1.Introduktion' on pages 2-4

The design begins by posing the main task to be investigated by the students

> T_1: 'How can we "best" estimate a parameter in a statistical model based on a sample?'

This task can also be formulated in terms of the example about women's foot lengths to evoke $\mathscr{P}_{Fisher}$ and $\mathscr{P}_{Context}$, respectively.

> $\mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context} :=$ [T_1: 'How can we "best" estimate the average foot length of Danish women based on a sample?']

In the first lesson, MLE is not introduced immediately as the solution technique to the task T_1. Instead, the first lesson focuses on the empirical and theoretical background needed to pose and answer T_1. The subject-didactic investigation has shown that MLE both belongs to the branch of statistics and the branch of functions. Therefore, the praxeological model consists of probability theory and statistics denoted $\Theta^*_{Prob\ \&\ stat}$ and functions denoted Θ^*_{Func} as theoretical elements. It is mainly $\Theta^*_{Prob\ \&\ stat}$ that will be supported in this lesson.

After posing T_1, the students are asked to recall their prior knowledge about parameters and estimators in exercises 1.1 and 1.2 [Appendix A, p. 2]. As an activity, some students' answers may be presented in plenary, so the teacher can evaluate students' understanding of parameters and estimates. Misunderstanding these concepts, may cause difficulties when students later are asked to construct statistical models. The plenary discussion activity may be continued by the teacher initiating a conversation about 'What does it mean that an estimate is the "best"?' to explore the meaning of T_1, since "best" is a quite vague and non-mathematical description. The students may suggest that the best estimate for the mean in the normal distribution is the sample mean. The teacher can motivate T_1 by explaining that we want to support this intuition with mathematical derivations. Otherwise, students may interpret "best" to mean the most reliable estimate, foreshadowing what some of the following lessons are about. The teacher should see it as a good thing that the students are thinking in this direction and mention that they will investigate this later.

Secondly, the stochastic situation of women's foot lengths is explained on page 3. I choose that this subchapter *"1. Introduktion"* and the next subchapter *"2. Maksimum likelihood estimatet for normalfordelingen"* [Appendix A] should use women's foot lengths as a recurring example, making the topic more accessible for students by including this contextualization. By using this example, students should explore the "statistical model" concept and acquire the competencies A1-A3 with emphasis on (A2) 'explain how stochastic situations can be translated into mathematical symbolic language'. This is also addressed in exercises 1.3 and 1.4, where students, among other things, are asked to construct statistical models.

To construct the statistical model in exercise 1.4 [Appendix A, p. 4], students have to switch between different forms of representation from words and data to a formal statistical description. For this, students can use the "random variable" aspect with the notation $X \sim N(\mu, \sigma)$ or $X \sim bin(n, p)$. Students may possibly use the strategy of copying the formulation at the bottom of page 3 [Appendix A]. This is not a bad strategy and it will likely give the correct result. Assuming that $X_1, \ldots, X_{56}$ are independent and $N(\mu, \sigma)$-distributed, where μ is unknown. Then, X_i represents the score of student i in the sample for $i = 1, 2, ..., 56$. The issues arise later if the student is asked to construct a binomial model. Then students may construct $n = 56$ random variables not understanding the structural difference between the normal model and the binomial model, because the students just copied the formulation.

It may also be that the stochastic part of the Grundvorstellung "prior assumptions" causes difficulties for students. For example, a binomial model (with $x = 2$ and $n = 10$) might be described using $X \sim bin(10, 0.2)$ with a known proportion $p = 0.2$. The correct answer is using $X \sim bin(10, p)$ with an unknown parameter p. These difficulties may arise because students struggle to understand the difference between p and $\hat{p}$. This is why I choose that students start by recalling their prior knowledge about estimates and parameters. Otherwise, students may not realize that a parameter represents the population proportion and not the sample proportion [Batanero and Borovcnik, 2016]. This heuristic difficulty may stem from a more general misunderstanding about the relationship between the sample and the population. Some students believe that the representativeness of the sample gives a perfect one-to-one correspondence between the sample and the population, and therefore state that $\hat{p} = 0.2$ is the true parameter. This belief is known in the statistics education literature as the representativeness heuristic [Batanero and Borovcnik, 2016].

7.2.2 Lesson 2: MLE for the Normal Distribution

Refer to appendix A '2.Maksimum likelihood estimat for normalfordelingen' on pages 4-7.

This lesson treats T_1 in the case of the normal distribution and teaches the students how to apply the technique of MLE, which in the reference model is denoted τ_{MLE}. T_1 can be answered in many different ways, but since I limit myself to the Fisherian praxeology, $\mathscr{P}_{Fisher}$, this naturally restricts the number of possible techniques for solving the task. In order for students to apply the τ_{MLE}, they must first solve the subtask T_0.

T_0 : 'How is the likelihood function derived?'

The task T_0, is solved by deriving the likelihood function using the "density" aspect. The technique of using the probability density function (or the point mass probability) to derive the likelihood function is denoted $\tau_{Density}$ in the reference model. This is done by applying the technology of "reversing" the view on which variable is fixed and which is varied in the density function, we denote this technology by $\theta^*_{Reverse}$. Other technologies are also necessary, for example using the multiplication principle to derive the joint density and logarithmic rules to determine the log-likelihood function. I will not explicitly include all the technologies in the reference model. Thus, one path of the reference model is $[T_0/\tau_{Density}/\theta^*_{Reverse}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$, which is supported by probability theory and statistics denoted $\Theta^*_{Prob\ \&\ stat}$ and the branch of calculus and functions denoted Θ^*_{Func} as theoretical elements. In this lesson, students will not explore the interpretation of the likelihood function in details, but just investigate T_0 as a subtask in order to answer T_1.

The investigation of T_0 and T_1 can as an activity be conducted through a plenary presentation performed by the teacher, with inserted sessions of exercises. The exercises focus on training or identifying the technologies used in $\tau_{Density}$ and τ_{MLE}. In exercises 2.1, 2.2 and 2.4, students train the involved technologies such as taking the logarithm of a product or differentiating a sum ($\sum$). As said, many technologies are involved in τ_{MLE}, including solving the score equation $\ell'(\theta) = 0$. I will denote this particular technology by θ^*_{Score}. In exercise 2.3, students are asked to identify the technologies involved in τ_{MLE} by explaining the equality signs in the derivation of the log-likelihood function, this exercise should be the main focus of lesson 2. The students can identify the technologies as follows. In showing that $\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) = -\frac{1}{2}\log\left(2\pi\sigma^2\right)$, one can write

$$\begin{aligned}\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) &\overset{(i)}{=} \log(1) - \log\left(\sqrt{2\pi\sigma^2}\right) \overset{(ii)}{=} -\log\left(\sqrt{2\pi\sigma^2}\right)\\ &\overset{(iii)}{=} -\log\left(\left(2\pi\sigma^2\right)^{\frac{1}{2}}\right) \overset{(iv)}{=} -\frac{1}{2}\log\left(2\pi\sigma^2\right).\end{aligned}$$

The technologies used here are (i) $\log(\frac{a}{b}) = \log(a) - \log(b)$, (ii) $\log(1) = 0$, (iii) $\sqrt{x} = x^{\frac{1}{2}}$ and (iv) $\log(x^a) = a \cdot \log(x)$.

Similar, when showing $-\sum_{i=1}^n \frac{1}{2}(\frac{x_i-\mu}{\sigma})^2 = -\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2$, one can write

$$-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2 \overset{(v)}{=} -\sum_{i=1}^{n}\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2} \overset{(vi)}{=} -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Here, the technologies are (v) $(\frac{a}{b})^n = \frac{a^n}{b^n}$ and (vi) $\sum_i a \cdot x_i = a\sum_i x_i$. Besides exploring the technologies used in T_0 and T_1, this exercise is also good for repetition of standard calculation rules, since a lot of such "standard" technologies are involved in the derivations.

7.2.3 Lesson 3: MLE for the Binomial Distribution

Refer to appendix A '3.Maksimum likelihood estimatet for binomialfordelingen' on pages 8-11.

Students are not expected to read the pages 8-11 at home. Considerations has been made regarding how the textbook chapter can guide the students to work more independently with τ_{MLE} when answering T_1. For that reason, I have chosen to structure this lesson on MLE for the binomial distribution in a quite different way than the preceding lesson on the normal distribution. The third lesson is mostly based on exercises that guide students through their own inquiry of deriving the maximum likelihood estimate. The lesson guides students through the path $[T_1/\tau_{MLE}/\theta^*_{Score}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$ of the praxeological model. In the exercises (3.1-3.11), students are required to apply the technologies that they identified in the previous lesson.

In the beginning of the third lesson, the example with opinion polls and the Danish Social Democratic Party is explained [Appendix A].

> $\mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context}$:=[T_1: 'How can we "best" estimate the probability of voting for the Social Democratic Party?']

Afterwards, the students are expected to work in small groups following the instructions of the textbook chapter. The first two exercises, 3.1 and 3.2, are a warm-up on the statistical model concept. The statistical model describing votes for the Social Democratic Party is provided in the chapter. Hence, students are not required to construct the model themselves. Instead, the lesson practices A1, which was not addressed in details in the first lesson. In exercise 3.1, students are asked to identify the stochastic part (the variability) of the model by asking whether two opinion polls would be exactly the same. In exercise 3.2, they are asked to identify the structural part of the model.

After the warm-up, the instructions in exercises 3.3-3.5 explore the task T_0. Since the students have only just been introduced to the technique $\tau_{Density}$, I have considered how much guidance the students should get when solving T_0. For example, the students are told that they should insert the specific values $n = 1635$ and $x = 322$ into the likelihood function, and that they should finally arrive at the log-likelihood function:

$$\ell_{322}(p) = \log\left(\binom{1635}{322}\right) + 322 \cdot \log(p) + 1313 \cdot \log(1-p).$$

The final hint is provided to ensure that students do not proceed with the incorrect log-likelihood function in the following exercises.

The next exercises focus on τ_{MLE}, where students estimate the probability parameter using MLE. In these exercises, students can develop the competencies E1–E3. The instructions also aim to introduce students to the likelihood intuition, B1, of 'comparability' or 'ranking'. In addition, the students should learn that the interpretation of the likelihood is the probability of observing the sample given a specific estimate. This interpretation is presented in the blue box on page 11 [Appendix A]. In exercise 3.11, students are asked to calculate the likelihood for some given estimates and rank them in order to train this intuition. Likewise, some exercises on the likelihood curve are included, both to support the intuition of ”ranking” by graphically representation and for students to recognize that the likelihood function and the log-likelihood function have their global maximum for the same value of p.

After students own guided inquiry, some groups may present their findings in plenary. At this point, the teacher should evaluate whether the students have grasped B1 and whether they interpret likelihoods correctly. Be aware that students may have done the exercises thoroughly and with a high learning outcome and at the same time have misinterpreted what a likelihood is. The correct interpretation of a likelihood is, as mentioned, the probability that the observed sample ($Data$) occurs given a specific estimate θ, which is denoted as $P(Data|\theta)$. The probability of the estimate given the observed sample is denoted $P(\theta|Data)$ and is an inverse probability/posterior distribution in Bayes' theorem. Believing that $P(Data|\theta) = P(\theta|Data)$ is a heuristic difficulty that Fisher made himself when he first introduced his ideas about MLE in 1912 [Edwards, 1974]. Fisher used Bayes' phrase of inverse probability in an incorrect way by confusing it with a likelihood. He claims that the maximum likelihood estimate was the most probable value for the parameter θ. This is wrong since the maximum likelihood estimator is the estimate making the observed data most likely [Pawitan, 2001c]. As said he made the mistake of reversing the two terms. Since then, the heuristic error of reversing conditions $P(\theta|Data) = P(Data|\theta)$ has proven to be one of the most pervasive mistakes in statistical terminology [Edwards, 1974]. The mistake does not lie in swapping the two probabilities in calculation, but in the interpretation. Fisher later, in 1922, specified the method for MLE in terms of likelihood functions and admitted his mistake from 1912 [Edwards, 1974].

Based on the heuristic error of reversing conditions, students may hypothetically answer exercise 3.7 by writing: ”The maximum likelihood estimate is the best estimate because it is the most probable value”. Also, students may answer exercise 3.11 by writing: ”Since $L_{322}(0.2) = 0.0236$, we can conclude that the probability, that the hypothesis $\hat{p_2} = 0.2$ is true, is 0.0236”. These errors advocate how wrong it can go on an interpretative level.

This difficulty was also spotted in the Japanese investigation [Hakamata et al., 2022] on the Brousseaunian guessing game.

> *"We have seen that the students naïvely replaced the (conditional) probability—which is, de facto, the likelihood in this context—with its inverse probability. Teachers tend to feel guilty when they overlook misunderstandings of their students. Indeed, mistakes are usually regarded as antididactic events under the paradigm of visiting works. By contrast, under the paradigm of questioning the world, any possible flaw involved in knowledge needs not always be avoided in advance; in fact, it can be welcomed in inquiry as probably didactic events."* [Hakamata et al., 2022, p. 11]

Even though my study does not work under the paradigm of questioning the world, I would still suggest, that as a teacher you let your students make the mistake as a didactic event. The teacher can then explain to the students that a likelihood is not a probability by showing with an example that the area under the likelihood curve is not 1, just as I did in the subject matter didactic analysis under the Grundvorstellung "reversed density" (6.2).

If we divide the likelihood function with the area under its curve, then we get a function that integrates to 1. This scaling of the likelihood function is called the *normalized likelihood* [Pawitan, 2001b]. Using the opinion poll example in lesson 3 the normalized likelihood function is

$$L_{322}^{norm}(p) = \frac{L_{322}(p)}{\int_0^1 L_{322}(p)dp}.$$

The interpretation of the normalized likelihood depends on whether you use a Bayesian or Fisherian view. In Bayesian statistics, the normalized likelihood will just be the inverse probability/posterior density when the axiomatic prior is uniform [Pawitan, 2001b]. In Fisherian statistics, axiomatic priors are not used, and the normalized likelihood can only be interpreted in terms of a likelihood even though it integrates to 1 [Pawitan, 2001c].

The question is then why the heuristic error of reversing conditions occurs in the interpretation of likelihood. The classic case, where two events A and B are confused such that the conditional probability $P(A|B)$ is mistaken for its inverse $P(B|A)$, is well-documented in the statistics education literature [Sotos et al., 2007]. It is often explained by students' misunderstanding of what 'A given B' actually means [Sotos et al., 2007]. However, this is not quite the same as confusing $P(\theta|Data)$ with $P(Data|\theta)$. In the setting of MLE, we have observed some data, so it seems logical to think of data as 'given'. Likewise, the object of interest is the estimate and hence it seems logical to interpret the likelihood as the probability of an estimate given data. The error may stems from students not understanding what is stochastic and what is not.

7.2.4 Lesson 4: Comparing Two Hypothesis

Refer to appendix A '4. Sammenligning af to hypoteser' on pages 12-16.

The forth lesson corresponds to pages 12–16, which the students are expected to have read at home. The last two lessons (4 and 5) treat the Grundvorstellung "relative measure", which gives rise the following tasks.

> T_2: 'Which of the two hypotheses $\hat{\theta}_1$ and $\hat{\theta}_2$ best describes the observed data?'
>
> T_3: 'How can we estimate the parameter θ reliably?'

Students can solve T_2 and T_3 by the technique of comparing two hypothesis $\hat{\theta}_1$ and $\hat{\theta}_2$ by their likelihoods. This technique is denoted by τ_{RM} (*relative measure*). The fourth lesson focuses on T_2, while the fifth lesson focuses on T_3.

This lesson starts by posing T_2 and introducing the concept of hypotheses, thereby letting students recall their prior knowledge from the second year course on the binomial test. The goal is to recall students' understanding that a hypothesis can be viewed as an estimate. The chapter presents the technologies needed to compare and rank hypotheses by their likelihoods, and it uses a recurring dice example to show how the technologies are applied. The dice example consists of a binomial model with $n = 10$, $x = 7$ and unknown probability parameter p [Appendix A, p. 13]. The first technology that students can apply is a direct comparison of the likelihoods using the inequality sign. This technology is denoted by $\theta^*_{Ineq.}$.

If $\hat{p}_1 = \frac{5}{6}$ and $\hat{p}_2 = \frac{1}{6}$, then student can use the calculation,

$$L_7(\hat{p}_1) = \binom{10}{7} \cdot \left(\frac{5}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^3 = 0.155$$

$$L_7(\hat{p}_2) = \binom{10}{7} \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^3 = 0.00025$$

to argue which hypothesis that best explain data $x = 7$ by $\theta^*_{Ineq.}$: $L_7(\hat{p}_1) > L_7(\hat{p}_2)$. Secondly, students can apply the technology of the likelihood ratio, $LR = \frac{L_x(\hat{p}_1)}{L_x(\hat{p}_2)}$, which is denoted θ^*_{LR}. The students can use the calculation,

$$LR = \frac{L_7(\hat{p}_1)}{L_7(\hat{p}_2)} = \frac{\binom{10}{7} \cdot \left(\frac{5}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^3}{\binom{10}{7} \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^3} = \frac{\left(\frac{5}{6}\right)^4}{\left(\frac{1}{6}\right)^4} = \frac{\frac{5^4}{6^4}}{\frac{1^4}{6^4}} = 5^4 = 620$$

to state that the data is 620 times more likely to occur under $\hat{p}_1$ than under $\hat{p}_2$.

The teacher must be aware that the likelihood ratio is not to be understood in terms of inverse probability. For example, students may conclude that $\hat{p}_1$ is 620 times more probable than $\hat{p}_2$, which is not correct.

Another difficulty can be that $\hat{p}_1$ is misunderstood as the deterministic true hypothesis, while $\hat{p}_2$ is misunderstood as the false hypothesis, because of $L_7(\hat{p}_1) > L_7(\hat{p}_2)$. This misunderstanding may come from the topic of binomial tests, where hypotheses are either rejected or not rejected, which can give the misconception that one hypothesis is always true and the other is then false. In a study of misconceptions in statistical interference [Sotos et al., 2007], it is explained how the misconception about deterministic false and true hypotheses comes from mathematical reasoning. Mathematical derivations are often associated with the proof that something is true. However, calculating the likelihoods is not a mathematical proof that one hypothesis is true and the other is false. Students' mathematical reasoning can therefore be disruptive in relation to using statistical knowledge in decision-making processes and for developing statistical literacy [Sotos et al., 2007].

Subchapter 4.5 is an exercise regarding coin tosses, which is very similar to the dice example. Students must formulate and compare hypotheses by calculating the likelihood ratio. In the final part of the exercise, students must explain what the likelihood ratio tells us about the two hypotheses. Subchapter 4.6 is an exercise, which is based on the hypergeometric probability distribution and the Lincoln-Peterson method. In the subject matter didactic analysis, I concluded that the maximum likelihood estimate for the hypergeometric likelihood function could not be derived by solving the score equation θ^*_{Score}. Therefore, I have chosen a graphical approach, where students must read the maximum likelihood estimate on the likelihood curve and graphically compare it to other hypotheses. The graphical approach gives rise to a third technology that can be used to compare hypotheses, namely by visual inspection, which is denoted θ^*_{VI} in my reference model. The visual inspection can support students' understanding of likelihood as a "relative measure" according to Kirsch's paradigm of 'accessibility by changing the mode of representation'. The task also has the subject-didactical purpose of emphasizing the generality of the likelihood approach by showing students how the techniques and technologies can be extended to other distributions. When students use θ^*_{VI} the teacher should emphasize that it is not the absolute distance between the heights on the likelihood curve, but rather the relative distance (the ratio) that make sense. The path of the praxeological model described in this lesson is $[T_2/\tau_{RM}/\theta^*_{Ineq.}, \theta^*_{LR}, \theta^*_{VI}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$.

7.2.5 Lesson 5: The Effect of Sample Size n

Refer to appendix A '5. Betydningen af stikprøvestørrelsen n' on pages 17-20.

The last lesson investigates the task

> T_3: 'How can we estimate the parameter θ reliably?'

The purpose of this lesson is for students to achieve the competence (E3) recognizing that the reliability of the maximum likelihood estimate depends on the sample size n. The lesson builds on an example with coin tosses to explain the effect of the sample size. It is not part of my design, that students perform the coin tosses themselves, since the effect of n should be explored by using the Fisherian praxeology $\mathscr{P}_{Fisher}$ and not by an informal approach.

Based on subchapter 5.1 [Appendix A, p. 17], the teacher poses T_3 and explains the example with coin tosses, where two binomial experiments are performed. One experiment with $n = 10$ coin tosses and one with $n = 50$, and where 'success' is getting a head.

> $\mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context} := [T_3$: 'How can we estimate the probability p of getting head reliably?']

In the first experiment, we observe $x = 8$ and in the second experiment $x = 40$. To investigate T_3, students must use the likelihood approach and apply the techniques they know from this course, τ_{MLE} and τ_{RM}. The students may try τ_{MLE} to determine the maximum likelihood estimate for both examples. The students can apply the technologies $\theta^*_{Reverse}$ and θ^*_{Score} or directly apply the formula $\hat{p}_{MLE} = \frac{x}{n}$. Students may recognize that the sample size do not effect the maximum likelihood estimate, since $\frac{40}{50} = \frac{8}{10} = 0.8$, and move on to the second technique τ_{RM}. To use τ_{RM}, student must formulate hypotheses, for example $\hat{p}_1 = 0.75$ and $\hat{p}_2 = 0.5$, determine $L_8(\hat{p}_1)$, $L_8(\hat{p}_2)$, $L_{40}(\hat{p}_1)$ and $L_{40}(\hat{p}_2)$ and compare the likelihoods.

$$\begin{aligned}
L_8(0.75) &= \binom{10}{8} \cdot 0.75^8 \cdot (1 - 0.75)^2 = 0.30199, \\
L_8(0.5) &= \binom{10}{8} \cdot 0.5^8 \cdot (1 - 0.5)^2 = 0.04394, \\
L_{40}(0.75) &= \binom{50}{40} \cdot 0.75^{40} \cdot (1 - 0.75)^{10} = 0.13982, \\
L_{40}(0.5) &= \binom{50}{40} \cdot 0.5^{40} \cdot (1 - 0.5)^{10} = 0.000009.
\end{aligned}$$

The students may now conclude that $L_8(0.75)$ is not the same as $L_{40}(0.75)$ indicating that the likelihood function is effected by the sample size n. It may be misunderstood that the hypothesis $\hat{\theta}_1 = 0.75$ is best supported by the small sample since $L_8(0.75) > L_{40}(0.75)$, but one cannot compare likelihood functions from different sample sizes.

Using τ_{RM} correctly, students should apply technologies θ^*_{VI} or θ^*_{LR} to compare the likelihoods. Using the technology θ^*_{LR}, students can calculate $LR_{10} = \frac{L_8(\hat{p}_1)}{L_8(\hat{p}_2)} = 6.41$ and $LR_{50} = \frac{L_{40}(\hat{p}_1)}{L_{40}(\hat{p}_2)} = 10798.18$. Then they should recognize that the statistical evidence for $\hat{p}_1 = 0.75$ compared to $\hat{p}_2 = 0.5$ becomes higher for increasing n.

With regards to the technology θ^*_{VI}, plotting the scaled likelihood functions with CAS-tools does not seem to strengthen the understanding of the topic, therefore this should be facilitated by the teacher. From the plot (Figure 6 in the chapter) the students can describe the shape of the two likelihood curves. Further, the teacher can ask the students what they think the likelihood curves for $n = 100$ with $x = 80$ and $n = 5$ with $x = 4$ will look like.

Finally, students can sit in groups and work on exercise 5.1 and the exercise in subchapter 5.4. Both exercises relate to the understanding of the relationship between the likelihood function and the sample size. In exercise 5.1, students can either set up the equation $L_5(\hat{p}_1) = L_5(\hat{p}_2)$ or use θ^*_{LR} to set up the equation $\frac{L_5(\hat{p}_1)}{L_5(\hat{p}_2)} = 1$. Here, students must use the interpretation of a likelihood as a ”relative measure” to understand what it means that the sample is equally likely under two hypotheses.

In subchapter 5.4 there is a sub-exercise of the same nature, but the sample size must now be determined such that the data are sixteen times more likely under one of the hypotheses compared to the other. Here, students must set $LR = 16$ and isolate for n. The student should observe that for the likelihood ratio to increase from 8 to 16 the sample size must also increase from 3 to 4.

7.2.6 Overview: Praxeological Reference Model

The praxeological organization consisted of three main tasks T_1, T_2 and T_3. The organization start with the method of MLE by posing the task T_1. The findings from the subject matter didactic analysis, particularly through the Grundvorstellung ”relative measure”, plays a significant role in the design and serves as the foundation for posing the tasks T_2 and T_3. All three tasks depends on the ability to derive the likelihood function, which lead to the subtask T_0. The model is developed to organize the different pathways that the design is intended to pursue. As such, it offers a comprehensive overview of the design with orientation towards practice. The complete praxeological reference model of the didactical design is

$$\begin{aligned} \mathscr{P}_{ref} &= \mathscr{P}_{Fisher} \oplus \mathscr{P}_{Conext} \\ &= [T_0, T_1, T_2, T_3/\tau_{Density}, \tau_{MLE}, \tau_{RM}/\theta^*_{Reverse}, \theta^*_{Score}, \theta^*_{Ineq.}, \theta^*_{LR}, \theta^*_{VI}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]. \end{aligned}$$

The **first lesson** is not build on a specific path in the reference model, but serves as an introduction to the topic. Its primary purpose is to elaborate on the theoretical element $\Theta^*_{Prob\ \&\ stat}$ by introducing a general statistical model. The purpose is to include the perspective of ”generality” (G1-G2) in the didactical design.

The **second lesson** builds on the path $[T_1/\tau_{MLE}/\theta^*_{Score}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$, where students access τ_{MLE} by identifying technologies that they already know from mathematics. Students should recognize knowledge/technologies from the normal distribution, logarithms and differential calculus, which adds the "coherence and repetition" perspective [H1-H3] to the didactical design.

The **third lesson** also builds on the path $[T_1/\tau_{MLE}/\theta^*_{Score}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$, but now the binomial distribution is used as case. The goal is for the students to work more independently with achieving τ_{MLE} for the binomial distribution. Also, the goal is to strengthen the perspective of "local connections" with connection to the probability and binomial distribution topic in 2.g. Thereby building local connections between the concepts of 'probability' and 'likelihood' as well as between the 'normal distribution' and 'binomial distribution', so that students experience that the different topics in school statistics are related (I1). My design do not explicitly cover the connection to the least square method (I2), this can be further explored in additional exercises or student projects.

The **fourth lesson** builds on path $[T_2/\tau_{RM}/\theta^*_{Ineq.}, \theta^*_{LR}, \theta^*_{VI}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$ and the **fifth lesson** builds on path $[T_3/\tau_{RM}/\theta^*_{Ineq.}, \theta^*_{LR}, \theta^*_{VI}/\Theta^*_{Prob\ \&\ stat}, \Theta^*_{Func}]$. The purpose is to teach new techniques and technologies so that students can approach MLE more critically and thereby develop statistical literacy. Thus, the design includes the perspective of "relativity", which is framed in F1-F5.

In Figure 12, an overview of the praxeological reference model is given, which is based on the previous investigation in the a priori analysis.

$\mathscr{P}_{ref} = \mathscr{P}_{Fisher} \oplus \mathscr{P}_{Context}$

Fisher's likelihood approach and the real-world context of the exercises.

T_0	T_1	T_2	T_3
How is the likelihood function derived?	How can we "best" estimate a parameter in a statistical model?	Which of the two hypotheses $\hat{\theta}_1$ and $\hat{\theta}_2$ best describes data?	How can we estimate the parameter θ reliably?

$\tau_{Density}$	τ_{MLE}	τ_{RM}
Deriving the likelihood from the probability density function.	Calculating the maximum likelihood estimate.	Comparing $\hat{\theta}_1$ and $\hat{\theta}_2$ by their likelihoods or comparing $\hat{\theta}_1$ and $\hat{\theta}_2$ by their likelihood for different sample sizes n.

$\theta^*_{Reverse}$	θ^*_{Score}	$\theta^*_{Ineq.}$, θ^*_{LR}, θ^*_{VI}
$L_x(\theta) = f_\theta(x)$, $\ell_x(\theta) = \log(L_x(\theta))$	Solution to $\ell'_x(\theta) = 0$.	$L(\hat{\theta}_1) > L(\hat{\theta}_2)$ Likelihood ratio: $LR = \frac{L(\hat{\theta}_1)}{L(\hat{\theta}_2)}$. Visual inspection.

$\Theta^*_{Prop\ \&\ stat}$, Θ^*_{Func}

The branch of probability theory and statistics,
and the branch of function analysis and differential calculus.

Figure 12: Praxeological reference model [Appendix B].

8 Discussion

The idea of introducing upper secondary school students to likelihood functions and MLE is investigated in this thesis applying a subject matter didactic analysis and a didactical design. One of the strengths has been to thoroughly and comprehensive explore likelihood functions and MLE, given that it is a totally new didactic transposition in a upper secondary school context. The subjective didactic analysis complements the didactical design effectively. Whereas the former emphasizes the theoretical and disciplinary elaboration of the content, the didactical design contains a more student- and context-oriented perspective. Another strength of my study is that I not only examine how MLE extend the statistical content in school mathematics, but also how MLE strengthens the subject of mathematics more broadly and contributes to coherence (in the "coherence and repetition" perspective). For this, the subject-didactic approach is highly beneficial in understanding how concepts are connected and structured in scholarly knowledge, while the context-oriented view is necessary to gain insight into the organization of Danish mathematics education.

Instead of investigating how MLE strengthens school mathematics and creates coherence, and developing a didactical design, other approaches could have been taken and different research questions explored. Another approach investigating likelihood functions and MLE could have been based on a comparative study, comparing the classical "frequentist" approach, the Bayesian approach and the likelihood approach to statistical inference. In such a study, one could have examined the subject-didactic advantages and disadvantages of the three approaches in relation to upper secondary school mathematics — and investigated whether it might be beneficial to combine different approaches to statistical inference in teaching. Within the Danish context, I think it is particularly relevant to examine the likelihood approach in relation to the classical approach involving binomial tests and confidence intervals. Notice that today only binomial test is core content (and not confidence intervals) in the Danish national syllabus [Børne- og Undervisningsministeriet, 2024a]. Students can perform a binomial test using their CAS-tool, which provides students with the acceptance region. They can then determine whether the observed data falls within or outside this region and, on that basis, decide whether to reject the hypothesis as in [Knud Nissen, 2013]. Of course, teaching the binomial test can be approached in many different ways, but the negative impacts of statistics-as-magic/black-box is plausible [Pedersen and Jankvist, 2021]. It would have been interesting to investigate likelihood functions in comparison to the core statistical content to see whether the likelihood approach could serve as a more analog alternative. Based on my investigation, the Grundvorstellung "relative measure" and "solving $\ell'(\theta) = 0$" supports a more analog approach. However, a comparative study focusing on digital and analog technologies would have been essential to underpin such statement.

One criticism of my study is that the textbook chapter has not been tested, which would have given rise to empirical data. When, as in this study, there is neither much literature to support the likelihood approach nor an empirical study conducted, the investigation of student strategies and difficulties becomes somewhat limited. Overall this makes it difficult to interpret the results of the a priori analysis and to identify which practical considerations the study may be lacking. Although I do not have any empirical material available, I critically reflect on the limitations of introducing likelihood functions and MLE in upper secondary school education. In particular, I have assessed the topic's theoretical nature and have tried to be realistic about the high demands regarding students' prerequisites. One could have considered lowering expectations and organized a different type of material. For example a didactical design only deriving the maximum likelihood estimate for the binomial distribution. In that case, the normal distribution and the density function could have been avoided as a prerequisite. The part building on the Grundvorstellung "relative measure" comparing different hypothesis could also have being avoided. The proposed topic would then lower the prerequisites a lot, but would lose the generality of MLE (the "generality" perspective) and development of students' statistical literacy (in the "relativity" perspective). On that basis, the design would lose its potential to strengthen the statistical content in school mathematics and mostly become an exercise in differential calculus and thus strengthen the mathematical content and not the statistical content. It has been essential that my didactical design includes the "generality" perspective and the "relativity" perspective in answering my main research question, which involves the strengthening of the statistical content of school mathematics.

9 Conclusion

The main research question investigated in this thesis is *"How can likelihood functions and maximum likelihood estimation be introduced in Danish upper secondary school mathematics in a way that strengthens the content in statistics and establishes connections between concepts in statistics, probability and mathematics?"*.

This question has been answered by a subject matter didactic analysis and a didactical design, which have been used to organize a textbook chapter on likelihood functions and MLE. Both analyses have contributed to insights on introducing the topic. The analyses complement each other, since they have different purposes and objects of analysis. In the subject matter didactic analysis, I analyse the scientific topic, describing the content in terms of aspects and Grundvorstellungen. The analysis provide a broad overview of the scholarly knowledge and help to unpack the "statistical model", "likelihood function" and "MLE" concepts. Moreover, the analysis supports subject-didactic arguments for why one should teach likelihood functions and MLE. The didactical design contains an institutional analysis and an a priori analysis. The content is approached from a completely different angle by the didactic engineering methodology looking into design implementation, task design and institutional conditions. This is useful to address what is feasible in practice, considering the different actors involved such as teachers, students and the national syllabus. By the a priori analysis the statistical knowledge is transformed and organized in a didactical design targeting upper secondary school students using praxeological organization. I consider the comprehensive didactical investigation to be useful in addressing the research questions, since the topic itself is largely unexplored in an upper secondary school context.

1.a) What knowledge will students acquire when working with the concepts of statistical model, likelihood function and MLE?

I have identified a range of competencies that students can acquire. With regard to statistical models students may learn to construct a statistical model and recognize the stochastic and structural part of the model. Working with likelihood students may learn to interpret the likelihood function in terms of a density function with reversed variables. Furthermore, the investigation showed that in order to understand likelihood functions, students need competencies that support the Grundvorstellung ”relative measure,” since likelihoods can only be interpreted in comparison with other likelihoods. Students will in this process be acquainted with likelihood ratio and visual inspection of the likelihood curve. Finally, with regard to MLE students may learn to solve the score equation using differential calculus and basic function analysis.

2.a) How can likelihood functions and MLE strengthen the content in statistics in upper secondary school?

To legitimize MLE as knowledge to be taught, four subject-didactic perspectives has been identified. Three of these perspectives strengthen the content of statistics in school mathematics. Firstly, the ”relativity” perspective, as framed in F5, contributes to the development of students’ statistical literacy by encouraging them to apply statistical knowledge and make statements based on statistical evidence. Secondly, the ”generality” perspective strengthen the statistical content by engaging students with various types of stochastic situations and statistical tasks, and thereby supporting statistical reasoning. Thirdly, the ”local connection” perspective strengthens the statistical content by creating local links between different topics within statistics. Moreover, my discussion considers how the likelihood approach could serve as an analog alternative to the current digital approach (statistics-as-magic) to statistics.

2.b) How can likelihood functions and MLE establish connections between different branches of school mathematics in upper secondary school?

The ”coherence and repetition” perspective demonstrates how statistics and mathematics is connected within the likelihood approach. The likelihood approach extend on the core content of statistics, e.g. density functions and statistical inference. Furthermore, to derive the maximum likelihood estimate basic function analysis and differential calculus are repeated — creating a strong connection to mathematics core content. This connection is used to argue that the likelihood approach links the branches of statistics and functions, and that determining the maximum likelihood estimate can be viewed as a shared problem space.

3.a) What institutional constrains and conditions are encountered when you design a textbook chapter about likelihood functions and MLE?

I conclude that likelihood functions and MLE can be implemented in upper secondary education as a supplementary topic in line with the 2024 syllabus (A-level). However, the investigation of institutional conditions reveal the high demands on both teachers and students in terms of qualifications and prior knowledge. Based on this, it is suggested that the topic is positioned in the third year of upper secondary school, so that students are already familiar with the density function of the normal distribution. It is also suggested that the design should not be implemented if statistics in general has been de-emphasized. Although teachers' knowledge of likelihood functions has not been specifically examined, it is anticipated that, given mathematics teachers' general difficulties with teaching the current core content of statistics, similar challenges might also arise when teaching the likelihood approach.

4.a) How can a teaching material on MLE be organized in a textbook chapter?

The design is developed as five lessons. The first half focuses on MLE, while the second half focuses on comparing hypothesis in terms of likelihoods. The practice-oriented part is based on a traditional textbook structure with exercises. What makes it different, is the particular emphasis on contextualization. Students will first encounter MLE for the normal distribution in order to train and identify the technologies used in the MLE technique. After this, MLE for binomial distribution is introduced, where students are expected to derive the maximum likelihood estimate themselves. At this stage, the practical work involves students applying the technologies independently. In the second half, students are asked to use the likelihood function to compare hypotheses. In practice, students are intended to apply two technologies, visual inspection and the likelihood ratio, in order to investigate hypotheses and perform critical statistical analyses.

4.b) How can tasks be formulated about MLE in a way that aligns with Danish upper secondary students' prerequisites and academic level?

Contextualization has been used as a key element throughout the textbook chapter to motivate and make the very theoretical topic more accessible to students using Kirsch's paradigm of 'making accessible by including the surroundings of mathematics'. In relation to deriving the maximum likelihood estimate for the binomial distribution, detailed instructions are provided to ensure that students are not left alone to explore a new and challenging technique on their own. In the exercises comparing hypotheses and exploring the effect of the sample size n, the focus is not only on calculation but also on interpretation trying to develop students' statistical literacy. Even though I have adapted the design to the target group and navigated the limitations, the a priori analysis shows that the interpretation of likelihoods can lead to obstacles that can be difficult to overcome. When investigating the task design, I found that students may encounter two major difficulties: the heuristic error of reversing conditions and the misconception about deterministic false and true hypotheses.

Overall, my investigation has shown that a didactical design on likelihood functions and MLE can be organized for upper secondary school in a way that strengthen the statistical content and that creates a link between mathematics and statistics. However, the investigation also shows a number of difficulties that students may encounter and other limitations, which might challenge the practical implementation of the design.

References

[Aarhus Universitet, 2025] Aarhus Universitet (2025). Matematisk statistik. https://kursuskatalog.au.dk/da/course/130413/Matematisk-Statistik. Course description for `Mathematical Statistics`. Accessed: 2025-05-17.

[Agermose Jensen and Timm, 2014] Agermose Jensen, M. and Timm, U. (2014). *Matemat10k : statistik : matematik B- og A-niveau*. Frydenlund, Frederiksberg, 1. edition.

[Barbe Farre et al., 2005] Barbe Farre, J., Bosch, M., Espinoza, L., and Gascón, J. (2005). Didactic restrictions on the teacher's practice: The case of limits of functions in spanish high schools. *Educational Studies in Mathematics*, 59:235–268.

[Batanero and Borovcnik, 2016] Batanero, C. and Borovcnik, M. (2016). *Statistics and Probability in High School*. SensePublishers Rotterdam.

[Biehler, 1994] Biehler, R. (1994). Probabilistic thinking, statistical reasoning, and the search for causes: Do we need a probabilistic revolution after we have taught data analysis? In Davidson, R. and Swift, J. B., editors, *Proceedings of the Fourth International Conference on Teaching Statistics (ICOTS 4)*, Marrakech, Morocco.

[Bosch and Gascón, 2014] Bosch, M. and Gascón, J. (2014). Introduction to the anthropological theory of the didactic (atd). In Bikner-Ahsbahs, A. and Prediger, S., editors, *Networking of Theories as a Research Practice in Mathematics Education*, pages 67–83. Springer International Publishing.

[Brydensholt and Ebbesen, 2025a] Brydensholt, M. and Ebbesen, G. R. (2025a). *Lærebog i matematik A2 stx*. Systime.

[Brydensholt and Ebbesen, 2025b] Brydensholt, M. and Ebbesen, G. R. (2025b). *Lærebog i matematik A3 stx*. Systime.

[Burrill and Biehler, 2011] Burrill, G. and Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*, pages 57–69. Springer.

[Børne- og Undervisningsministeriet, 2024a] Børne- og Undervisningsministeriet (2024a). Matematik a – stx, august 2024.

[Børne- og Undervisningsministeriet, 2024b] Børne- og Undervisningsministeriet (2024b). Vejledning til læreplan i matematik a, stx august 2024.

[Børne- og Undervisningsministeriet, 2024c] Børne- og Undervisningsministeriet (2024c). Vejledning til studieretningsprojektet, stx. https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf23/aug/vejledninger/240807-studieretningsprojektet-stx.pdf. Accessed: 2025-04-29.

[Børne- og Undervisningsministeriet, 2024d] Børne- og Undervisningsministeriet (2024d). Vurdering af fagniveauer. https://www.uvm.dk/-/media/filer/uvm/udd/gym/pdf23/aug/vejledninger/240807-studieretningsprojektet-stx.pdf. Accessed: 2025-05-16.

[Carstensen, 2021] Carstensen, J. (2021). *Mat A2 stx*. Systime, Aarhus, 4. edition.

[Chevallard and Bosch, 2014] Chevallard, Y. and Bosch, M. (2014). *Didactic Transposition in Mathematics Education*, pages 170–174. Springer Netherlands, Dordrecht.

[Clausen et al., 1997] Clausen, F., Prinz, P., and Schomacker, G. (1997). *Sandsynlighedsregning og statistik*. Munksgaard.

[Cobb and Moore, 1997] Cobb, G. W. and Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9):801–823.

[Det Naturvidenskabelige Studienævn SDU, 2025] Det Naturvidenskabelige Studienævn SDU (2025). ST521: Matematisk statistik. https://odin.sdu.dk/sitecore/index.php?a=fagbesk&id=139752&lang=da. Syddansk Universitet. Accessed: 2025-05-17.

[Ditlevsen and Sørensen, 2018] Ditlevsen, S. and Sørensen, H. (2018). *Introduktion til Statistik*. Institut for Matematiske Fag, Københavns Universitet, 5. edition.

[Edwards, 1974] Edwards, A. W. F. (1974). The history of likelihood. *International Statistical Review / Revue Internationale de Statistique*, 42(1):9–15.

[Efron and Hastie, 2016] Efron, B. and Hastie, T. (2016). *Fisherian Inference and Maximum Likelihood Estimation*, pages 38–52. Institute of Mathematical Statistics Monographs. Cambridge University Press.

[Ekstrøm et al., 2017] Ekstrøm, C. T., Hansen, E., and Brockhoff, P. B. (2017). Statistik i gymnasiet. *LMFK-bladet*, (2):14–20.

[Etz, 2018] Etz, A. (2018). Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science*, 1(1):60–69.

[Fisher, 1922] Fisher, R. A. (1922). Inverse probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society*, 28:257–261.

[González-Martín et al., 2014] González-Martín, A. S., Bloch, I., Durand-Guerrier, V., and Maschietto, M. (2014). Didactic situations and didactical engineering in university mathematics: cases from the study of calculus and proof. *Research in Mathematics Education*, 16(2):117–134.

[Greefrath et al., 2016] Greefrath, G., Oldenburg, R., Siller, H.-S., Ulm, V., and Weigand, H.-G. (2016). Aspects and “grundvorstellungen” of the concepts of derivative and integral. *Journal für Mathematik-Didaktik*, 37:99–129.

[Grøn, 2015] Grøn, B. (2015). *Hvad er matematik? - B: grundbog.* L & R Uddannelse.

[Groth, 2015] Groth, R. E. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education*, 46(1):4–16.

[Gråbæk et al., 2025] Gråbæk, L., Schmidt, S. N., and Vestergaard, L. (2025). *Statistik C.* Systime.

[Hadi and Sahib, 2023] Hadi, N. and Sahib, A. (2023). Review of maximum likelihood estimation method. *International Journal of Engineering and Information Systems (IJEAIS)*, 7(8):20–23.

[Hakamata et al., 2022] Hakamata, R., Otaki, K., Fukuda, H., and Otani, H. (2022). Statistical modelling in the Brousseaunian guessing game: A case of teacher education in Japan. In *Fourth conference of the International Network for Didactic Research in University Mathematics*, Hannover, Germany. Leibnitz Universität.

[Harradine et al., 2011] Harradine, A., Batanero, C., and Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*, volume 14, pages 235–246. Springer.

[Hebsgaard and Sloth, 1999] Hebsgaard, T. and Sloth, H. (1999). *Højniveaumatematik. Bind 2*. Forlaget Systime, Aarhus, Denmark.

[Hofe and Blum, 2016] Hofe, R. and Blum, W. (2016). "grundvorstellungen" as a category of subject-matter didactics. *Journal für Mathematik-Didaktik*, 37:225–254.

[Hußmann et al., 2016] Hußmann, S., Rezat, S., and Sträßer, R. (2016). Subject matter didactics in mathematics education. *Journal für Mathematik-Didaktik*, 37:1–9.

[IASE, 2025] IASE (2025). Iase publications search. https://iase-web.org/search.php. Accessed: 2025-04-29.

[Jacod and Protter, 2012] Jacod, J. and Protter, P. (2012). *Probability Essentials*. Universitext. Springer Berlin Heidelberg.

[Kang and Kilpatrick, 1992] Kang, W. and Kilpatrick, J. (1992). Didactic transposition in mathematics textbooks. *For the learning of mathematics*, 12:2–7.

[Knud Nissen, 2013] Knud Nissen (2013). Maple 17 b-niveau. https://maplesoft.com/MapleGym/Maple17BNiveau_ebog.pdf. MapleGym. Accessed: 2025-05-31.

[Københavns Universitet, 2025] Københavns Universitet (2025). NMAA09014U – Sandsynlighedsregning og statistik (SS). https://kurser.ku.dk/course/nmaa09014u. Course description for Probability Theory and Statistics (SS). Accessed: 2025-05-17.

[Larsen and Jensen, 2019] Larsen, J. C. and Jensen, K. B. S. (2019). Metoder og videnskabsteori i, med og om matematik. https://emu.dk/sites/default/files/2019-11/Mat%20Metode%20-%20ENDELIG%20TEKST.pdf. Accessed: 2025-05-30.

[Lauritzen, 2023] Lauritzen, S. (2023). *Fundamentals of Mathematical Statistics*. CRC Press.

[Makar and Rubin, 2017] Makar, K. and Rubin, A. (2017). *Learning About Statistical Inference*, pages 261–294. Springer.

[Markussen, 2020] Markussen, B. (2020). Normalfordelingen. https://emu.dk/sites/default/files/2020-06/normalfordelingen.pdf. Børne- og Undervisningsministeriet. Accessed: 2025-05-16.

[Milhøj, 2025] Milhøj, A. (2025). Teoretisk statistik. https://lex.dk/teoretisk_statistik. Lex.dk, redigeret af Den Store Danske.

[Nielsen, 2017] Nielsen, H. B. (2017). *Introduction to Likelihood-based Estimation and Inference*. Hans Reitzel, 3 edition.

[Ottaviani, 2011] Ottaviani, M. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. In *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education*, volume 14, pages 121–132. Springer.

[Pawitan, 2001a] Pawitan, Y. (2001a). Basic models and simple applications. In *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, pages 73–115. Oxford University Press.

[Pawitan, 2001b] Pawitan, Y. (2001b). Elements of likelihood inference. In *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, pages 20–52. Oxford University Press.

[Pawitan, 2001c] Pawitan, Y. (2001c). Introduction. In *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, pages 1–19. Oxford University Press.

[Pedersen and Jankvist, 2021] Pedersen, M. K. and Jankvist, U. T. (2021). Cas som black box. https://matematikdidaktik.dk/tema/digitale-teknologier/cas-som-black-box. DIGITALE TEKNOLOGIER. Accessed: 2025-05-30.

[Reid, 2000] Reid, N. (2000). Likelihood. *Journal of the American Statistical Association*, 95(452):1335–1340.

[Scheaffer, 2006] Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. In Burrill, G. F., editor, *Thinking and Reasoning with Data and Chance: Sixty-eighth Yearbook*, pages 309–321. National Council of Teachers of Mathematics, Reston.

[Scheiner et al., 2023] Scheiner, T., Buchholtz, N., and Kaiser, G. (2023). Mathematical knowledge for teaching and mathematics didactic knowledge: a comparative study. *Journal of Mathematics Teacher Education*, 27:1083–1104.

[Schilling, 2017] Schilling, R. (2017). *Measures, Integrals and Martingales*. Measures, Integrals and Martingales. Cambridge University Press.

[Sotos et al., 2007] Sotos, A., Vanhoof, S., Van den Noortgate, W., and Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2:98–113.

[Tietze et al., 2002] Tietze, U.-P., Klika, M., and Wolpers, H. (2002). *Mathematikunterricht in der Sekundarstufe II: Band 3: Didaktik der Stochastik*. Vieweg+Teubner Verlag Wiesbaden.

[Varberg, 1963] Varberg, D. E. (1963). The development of modern statistics. *The Mathematics Teacher*, 56(4):252–257.

[Vohns, 2016] Vohns, A. (2016). Fundamental ideas as a guiding category in mathematics education—early understandings, developments in german-speaking countries and relations to subject matter didactics. *Journal für Mathematik-Didaktik*, 37:193 – 223.

[Watkins, 2011] Watkins, J. C. (2011). Topic 15: Maximum likelihood esitmation*. https://math.arizona.edu/~jwatkins/o-mle.pdf. The University of Arizona. Accessed: 2025-05-30. *(Lecture notes)*.

[Weiland, 2019] Weiland, T. (2019). Critical mathematics education and statistics education: Possibilities for transforming the school mathematics curriculum. In Chernoff, E. J.

and Sriraman, B. E., editors, *Topics and Trends in Current Statistics Education Research*, pages 391–411. Springer.

[Winsløw, 2011] Winsløw, C. (2011). Anthropological theory of didactic phenomena: Some examples and principles of its use in the study of mathematics education. *Un panorama de TAD. CRM Documents*, 10:117–138.

[Zieffler et al., 2017] Zieffler, A., Garfield, J., and Fry, E. (2017). *What Is Statistics Education?*, pages 37–70. Springer International Handbooks of Education.

A The Textbook Chapter on MLE

One the next page begins the textbook chapter about likelihood functions and MLE.

The chapter is developed by me for this thesis.

Jagten på det Bedste Estimat

En Introduktion til Likelihoodfunktioner og Maksimum Likelihood Estimatet (*Supplerende Emne*)

Kilde: "Isfugl, han". Fotografi af Peter Halkier. https://nfd.dk/billede/isfugl-han/.

Supplerende kapitel skrevet af Cæcilie Bøje Pedersen

Faktaboks om Fisher og Likelihoodfunktionen

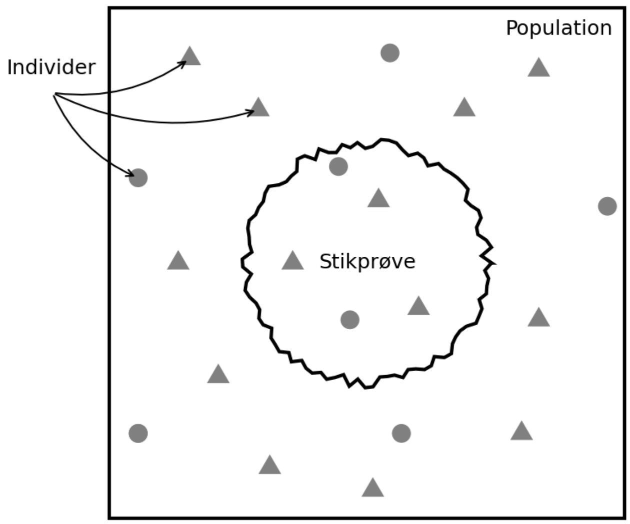
Likelihoodfunktionen og maksimum likelihood estimatet blev opfundet af **Sir Ronald Aylmer Fisher (1890-1962)**, som var en engelsk statistiker, der arbejdede med biologi og genetik. Fisher opfandt de to begreber samtidigt i perioden mellem 1912 og 1922.

I dag betragter man Fisher som en pioner indenfor statistikfaget, fordi han sammen med to andre statistikere Jerzy Neyman og Karl Pearson var med til at udvikle den videnskabelige måde vi i dag laver statistisk på ved bl.a. statistiske tests.

Fishers maksimum likelihood estimation er i dag en meget populær statistisk metode, som for eksempel bliver anvendt indenfor biologi, psykologi, medicin og datavidenskab.

1. Introduktion

I statistik indsamler vi ofte stikprøver med det formål at undersøge en bestemt egenskab ved en population. Vi vil som regel gerne finde ud af hvilken model og hvilke parametre, der bedst beskriver den observerede stikprøve for at kunne sige noget mere generelt om egenskaben hos populationen.



Figur 1: Illustration af population og stikprøve

En sådan statistisk undersøgelse kan omhandle mange forskellige fænomener fra verden omkring os. For eksempel kan vi undersøge menneskers højde, vælgertilslutningen til et politisk parti og mekanismen bag terningekast. For undersøgelser der tager udgangspunkt i en stikprøve, kan vi introducere begrebet *likelihoodfunktionen*.

Likelihoodfunktioner kan bruges til at finde den værdi for en parameter, som bedst forklarer variationen i stikprøven. Det kan for eksempel være, at vi ønsker at bestemme den bedst mulige værdi for middelværdiparameteren μ i en normalfordeling ud fra en observeret stikprøve $x = (x_1, x_2, \dots, x_n)$. Et andet eksempel kunne være at bestemme den bedst mulige værdi for sandsynlighedsparameteren p i en binomialfordeling ud fra en observeret stikprøve x (antal succeser i stikprøven). Vi kan i princippet gætte på, hvad værdierne af parametrene kan være på baggrund af stikprøven, men det anvendelige ved likelihoodfunktionen er, at vi kan bestemme værdierne på en systematisk måde ved matematiske udledninger. Dette bud kaldes *maksimum likelihood estimatet*.

I dette kapitel skal vi studere likelihoodfunktioner og se hvordan de kan anvendes til at bestemme maksimum likelihood estimater for ukendte parametre. Vi vil som udgangspunkt bruge normalfordelingen og binomialfordelingen som eksempel til at introducere de nye begreber i kapitlet.

Øvelse 1.1: Giv eksempler på hvad en parameter er fra de eksempler du tidligere har arbejdet med i matematikundervisningen?

Øvelse 1.2: Anvend internettet til at undersøge hvad ordene *estimere* og *estimat* betyder?

1.1 Population og stikprøve

I et statistisk forsøg ønsker vi at undersøge en egenskab for en population. For eksempel fodlængden hos voksne danske kvinder. Her er *populationen* voksne danske kvinder og *egenskaben* er kvindernes fodlængde.

Vi foretager undersøgelsen ved tilfældigt at udvælge en stikprøve på n individer fra populationen og så observere egenskaben hos individerne i stikprøven. Vi udtager en stikprøve, fordi man sjældent har ressourcer nok til at undersøge egenskaben i hele populationen.

I undersøgelsen udvælger vi 20 voksne kvinder og måler deres fodlængde i millimeter. Her er stikprøven givet i tabellen.

Stikprøve 1									
224	232	235	237	237	241	242	243	247	247
248	249	252	253	256	259	260	264	267	274

1.2 Den statistiske model

Vi opstiller en statistisk model, hvor X er en stokastisk variabel, der angiver egenskaben for et tilfældigt individ i populationen. Vi antager også, at egenskaben hos individerne er uafhængig af hinanden, hvilket ofte kun er delvist opfyldt. Uafhængigheden sikrer, at vi kan anvende *multiplikationsprincippet/både-og princippet*, hvilket bliver vigtigt senere.

Vi kan nu danne n uafhængige stokastiske variable, en for hvert individ (kvinde) i stikprøven, som alle er normalfordelt med ukendt middelværdiparameter μ. Vi vil generelt antage at spredningsparameteren σ er kendt på forhånd og derfor ikke bekymre os om denne parameter.
Vi kan da opstille følgende statistiske model.

Model: $X_1, X_2, \dots, X_{20}$ er uafhængige og $N(\mu, \sigma)$-fordelte, hvor μ er ukendt.
Her angiver X_i fodlængden for kvinde i i stikprøven for $i = 1, 2, \dots, 20$.

Hvad er en statistisk model?

Udgangspunktet i en statistisk undersøgelse er en statistisk model. Den statistiske model er en matematisk beskrivelse af et virkeligt fænomen, hvor udfaldet ikke er givet på forhånd.

Definition 1: En statistisk model består af en stokastisk variabel X, som har en sandsynligheds-fordeling med en ukendt parameter.

Øvelse 1.3:

a) Giv eksempler på virkelige fænomener, som kan modelleres ved en normalfordeling.

b) Giv også nogle eksempler på virkelige fænomener, som kan modelleres ved en binomial fordeling.

Øvelse 1.4:

Til matematikscreeningen i 1.g kan eleverne få mellem 0 og 100 point. Der udvælges tilfældigt en stikprøve på 56 elever fra 1.g, hvor man noterer deres pointtal.
Nedenunder ses den sorterede stikprøve i tabellen.

Stikprøve 2													
6	10	16	19	23	24	26	29	34	34	35	35	41	42
43	43	41	45	46	46	46	47	48	52	54	54	56	57
57	59	60	60	60	61	62	62	62	64	64	65	66	66
66	66	68	69	70	71	72	73	77	78	80	81	85	93

a) Angiv population og stikprøve i undersøgelsen?

b) Opstil en statistisk model, som beskriver pointfordelingen ved matematikscreeningen i 1.g.

2. Maksimum likelihood estimat for normalfordelingen

2.1. Den sande værdi og estimatet

Parameteren μ er et tal, der ikke kun beskriver stikprøven, men hele populationen. Ved at estimere denne parameter får vi derved ny viden om den population, som vi undersøger. I undersøgelsen om voksne kvinders fodlængde interesserer vi os for middelværdien af kvinders fodlængde for hele den kvindelige befolkning. Da vi som sagt ikke kan måle/observere hele populationen vil μ være ukendt. Vi vil kalde den ukendte værdi for den sande værdi af μ. De bud man kan give på værdien af μ kaldes for estimater og benævnes med $\hat{\mu}$. Så nu går jagten ind på at bestemme estimatet for μ.

2.2 Maksimum Likelihood metoden

Maksimum likelihood metoden går ud på at estimere parametrene i en antaget statistisk model, på baggrund af en stikprøve. Metoden bygger på antagelsen, at det vi faktisk har observeret, må være det, der er mest sandsynligt at observere.

En normalfordeling med middelværdiparameter μ og spredning σ har tæthedsfunktion

$$f_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} \ .$$

Spredningen σ betragtes som en kendt konstant og variablen x er et vilkårligt udfald fra udfaldsrummet. *Likelihoodfunktionen* er identisk med tæthedsfunktionen, men parameteren μ betragtes som en variabel mens observationen x betragtes som en konstant: altså er der byttet rundt på deres rolle.

Likelihood funktionen for en observation x benævnes

$$L_x(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} \ .$$

Maksimum likelihood metoden går ud på, at vi skal bestemme det globale maksimumssted for likelihoodfunktionen L_x. Vi leder altså efter det μ, hvor likelihoodfunktionen er størst mulig.

Når stikprøven består af flere observationer $x_1, x_2, \ldots, x_n$ benævnes likelihoodfunktionen $L_{x_1, x_2, \ldots, x_n}(\mu)$. Da $X_1, X_2, \ldots, X_n$ er uafhængige kan vi bruge multiplikationsprincippet til at udlede likelihoodfunktionen for samtlige observationer.

$$L_{x_1, x_2, \ldots, x_n}(\mu) = f_\mu(x_1) \cdot f_\mu(x_2) \cdot \ldots \cdot f_\mu(x_n) = \prod_{i=1}^{n} f_\mu(x_i).$$

Øvelse 2.1:

a) Skriv summerne og produktet ud.

$$i)\ \sum_{i=1}^{4} x_i \qquad ii)\ \prod_{i=1}^{4} x_i \qquad iii)\ \sum_{i=1}^{4} \frac{1}{2} x_i$$

$$iv)\ \sum_{i=1}^{4} \frac{1}{\sqrt{2\pi\sigma^2}} x_i \qquad v)\ \sum_{i=1}^{4} 3\ .$$

b) Vi har indsamlet en stikprøve $x_1 = 10$, $x_2 = 12$, $x_3 = 10$ og $x_4 = 11$.

Udregn de fem udtryk ud fra stikprøven.

Log-likelihoodfunktionen

Da den naturlige logaritme er en voksende funktion, gælder der, at $L_x(\mu)$ og $\ln(L_x(\mu))$ har maksimum for samme værdi af μ. Dette vil vi ikke bevise, men det er et meget brugbart resultat, da det er meget nemmere at differentiere $\ln(L_x(\mu))$ end $L_x(\mu)$.

Funktionen $\ln(L_x(\mu))$ kalder vi *log-likelihoodfunktionen* og benævnes $l_x(\mu)$.

Givet observationerne $x_1, x_2, \dots, x_n$ bestemmes log-likelihoodfunktionen $l_{x_1,\, x_2, \dots,\, x_n}(\mu)$ ved at tage den naturlige logaritme af likelihoodfunktionen $\prod_{i=1}^{n} f_\mu(x_i)$.

$$l_{x_1,\, x_2, \dots,\, x_n}(\mu) = \ln\left(\prod_{i=1}^{n} f_\mu(x_i)\right)$$

Vi kan da omforme udtrykket ved brug af logaritmeregnereglerne. Tager vi logaritmen af et produkt bliver det til en sum af logaritmer.

$$\ln\left(\prod_{i=1}^{n} f_\mu(x_i)\right) = \sum_{i=1}^{n} \ln\left(f_\mu(x_i)\right) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\cdot\left(\frac{x_i-\mu}{\sigma}\right)^2}\right) =$$

$$\sum_{i=1}^{n}\left(ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2}\cdot\left(\frac{x_i-\mu}{\sigma}\right)^2\right) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2}\cdot\left(\frac{x_i-\mu}{\sigma}\right)^2\right) =$$

$$-\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{1}{2}\cdot\left(\frac{x_i-\mu}{\sigma}\right)^2 = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Dvs. log-likelihoodfunktionen er $l_{x_1,\, x_2, \dots,\, x_n}(\mu\,) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$.

Øvelse 2.2: Vi har en stikprøve $x_1 = 22$, $x_2 = 24$, $x_3 = 25$, $x_4 = 24$ og $x_5 = 20$.
Beregn følgende udtryk ud fra stikprøven.

$$a)\ \ln\left(\prod_{i=1}^{5} x_i\right) \qquad b)\ ln(2\cdot e^{x_1}) \qquad c)\ \sum_{i=1}^{5} ln(2\cdot e^{x_1}) \qquad d)\ \sum_{i=1}^{5}\mu.$$

Øvelse 2.3: Forklar lighedstegnene i udledningen af log-likelihoodfunktionen ovenfor.

2.3 Maksimering af log-likelihood funktionen

Den afledte funktion af $l_{x_1, x_2,\dots, x_n}(\mu)$ bliver en relativt pæn funktion

$$l_{x_1, x_2,\dots, x_n}{}'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu),$$

da det første led i log-likelihoodfunktionen giver nul, da leddet ikke afhænger af parameteren μ.

Øvelse 2.4: Differentier følgende udtryk i forhold til μ.

$$a)\ \sum_{i=1}^{6}(\mu - i)^2 \qquad\qquad b)\ \ln(2\pi\sigma^2) + \sum_{i=1}^{6}(\mu - i)^2$$

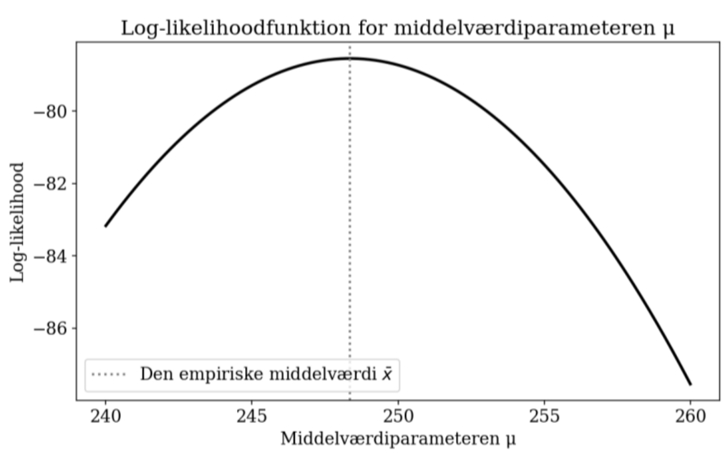
Vi bestemmer nu de punkter på grafen for log-likelihoodfunktionen $l_{x_1, x_2,\dots, x_n}(\mu)$ med vandret tangent ved at sætte den afledte lig med nul.

$$l_{x_1, x_2,\dots, x_n}{}'(\mu) = 0 \Leftrightarrow$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \ \Leftrightarrow\ \sum_{i=1}^{n}(x_i - \mu) = 0 \ \Leftrightarrow\ \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu = 0 \ \Leftrightarrow$$

$$\sum_{i=1}^{n}x_i - n\cdot\mu = 0 \ \Leftrightarrow\ \sum_{i=1}^{n}x_i = n\cdot\mu \ \Leftrightarrow\ \mu = \frac{1}{n}\sum_{i=1}^{n}x_i \ \Leftrightarrow\ \mu = \bar{x}$$

Dvs. der er en vandret tangent ved $\mu = \bar{x}$, hvor $\bar{x}$ angiver gennemsnittet af observationerne.
Nedenfor ses grafen for log-likelihoodfunktionen. Det fremgår at funktionen har et globalt maksimum ved $\mu = \bar{x}$, som er angivet med den stiplet linje i figuren.



Figur 2: Log-likelihoodfunktionen for middelværdiparameteren

Det vil sige maksimum likelihood estimatet for μ i normalfordelingen er $\hat{\mu} = \bar{x}$.

2.4 Beregning af estimatet

Vi vender nu tilbage til eksemplet med danske kvinders fodlængde, hvor stikprøvestørrelsen er $n = 20$. Maksimum likelihood estimatet $\hat{\mu}$ for danske kvinders fodlængde bliver

$$\hat{\mu} = \bar{x} = \frac{224 + 232 + 235 + \cdots + 264 + 267 + 274}{20} = 248{,}4.$$

Vi konkluderer, at danske kvinder har en gennemsnitlig fodlængde på $248{,}4$ millimeter. Bemærk at udtalelsen ikke kun gælder kvinderne i stikprøven, men danske kvinder generelt.

Øvelse 2.5: Udregn maksimum likelihood estimatet for stikprøven i *øvelse 2.2* med CAS-værktøj.

3. Maksimum likelihood estimat for binomialfordelingen

3.1 Vælgertilslutningen til Socialdemokratiet

En statistisk undersøgelse omhandler vælgertilslutningen til det danske parti Socialdemokratiet, hvis der var folketingsvalg i morgen. Her er populationen stemmeberettigede danskere og egenskaben er om man stemmer socialdemokratisk eller ej. I undersøgelsen vil vi bruge en binomialmodel til at beskrive fordelingen af socialdemokratiske partivalg.

Øvelse 3.1: Lav en søgning på internettet, hvor I søger efter politiske meningsmålinger. Hvad er antallet af socialdemokratiske partivalg ifølge meningsmålingen? Viser to forskellige meningsmålinger præcis den samme vælgertilslutning til Socialdemokratiet – hvor er de forskellige?

Danmarks Radio (DR) laver månedlige meningsmålinger baseret på interviews, hvor de spørger hvad personen ville stemme på, hvis der var folketingsvalg i morgen. I perioden d. 15.-22. januar 2025 afgav 1635 stemmeberettede danskere et partivalg til DR. Ud af de 1635 interviewede sagde $x = 322$ danskere, at de ville stemme på Socialdemokratiet.

Statistisk model: Lad X være antallet af socialdemokratiske partivalg i en stikprøve på $n = 1635$ stemmeberettigede danskere. X er $bin(1635, p)$-fordelt med antalsparameter $n = 1635$ og ukendt sandsynlighedsparameter p.

Øvelse 3.2: Overvej under hvilke antagelser X er binomialfordelt. Inddrag begreberne ”med tilbagelægning” og ”uden tilbagelægning”.

Vi har tidligere i undervisningen ladet estimatet for sandsynlighedsparameteren p i en binomialfordeling være frekvensen x/n, hvilket også giver intuitiv god mening, når det er en sandsynlighed vi estimerer. I dette afsnit skal vi bevise at maksimum likelihood estimatet for sandsynlighedsparameteren p i en binomialfordeling netop er $\hat{p} = \frac{x}{n} = \frac{322}{1635} = 0{,}197$.

3.2 Log-likelihoodfunktionen for binomialfordelingen

Punktsandsynlighederne for en binomialfordeling er givet ved

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x},$$

hvor p er sandsynlighedsparameteren og n er antalsparameteren.

Øvelse 3.3: Opskriv forskriften for likelihoodfunktionen, som er identisk med sandsynlighedsfunktionen ovenfor, men hvor vi betragter p som den uafhængige variable og x som en fast værdi.

$\boldsymbol{L_x(p) = \ldots}$

Indsæt værdierne $n = 1635$ og $x = 322$ i likelihoodfunktionen og opskriv $\mathrm{L}_{322}(\mathrm{p})$.

$\boldsymbol{L_{322}(p) = \ldots}$

Øvelse 3.4: Det er en god ide at tage logaritmen af likelihoodfunktionen $L_{322}(p)$. Denne funktion kaldes som sagt for log-likelihoodfunktionen. Bestem $l_{322}(p)$.

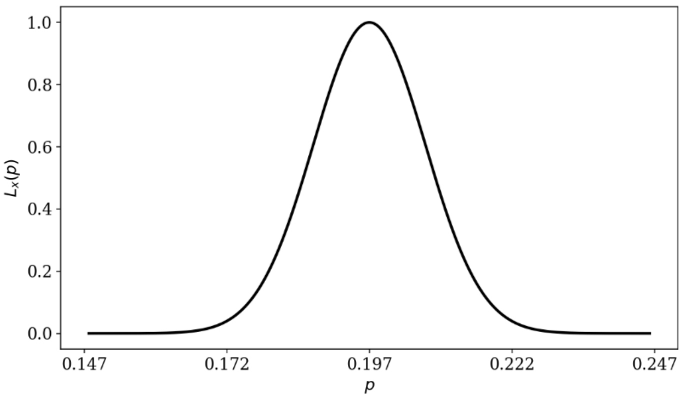
$\boldsymbol{l_{322}(p) = \ldots}$

Øvelse 3.5: Brug logaritmeregnereglerne til at reducere funktionsudtrykket for $l_{322}(p)$ og vis at $l_{322}(p) = \ln\left(\binom{1635}{322}\right) + 322 \cdot \ln(p) + 1313 \cdot \ln(1-p)$

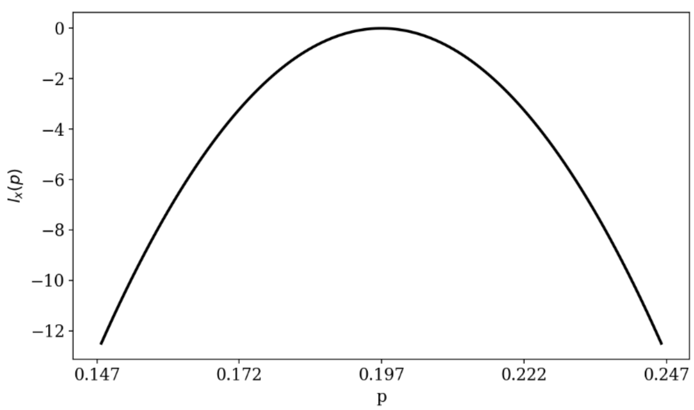
(Se formelsamling for logaritmeregneregler.)

3.3 Sammenligning af likelihoodfunktionen og log-likelihoodfunktionen

Her ses grafen for likelihoodfunktionen $\mathrm{L}_{322}(\mathrm{p})$ og log-likelihoodfunktionen $l_{322}(\mathrm{p})$ i hver deres plot. I de to plots er funktionerne skaleret så værdien af det globale maksimum er hhv. 1 og 0.



Figur 3: Likelihoodfunktionen $\boldsymbol{L_{322}(p)}$



Figur 4: Log- likelihoodfunktionen $\boldsymbol{l_{322}(p)}$

Øvelse 3.7:

Forklar hvorfor likelihoodfunktionens maksimumssted kan anses for at være det bedste bud på et estimat for binomialmodellen.

Øvelse 3.8:

Aflæs det sted hvor grafen for $L_{322}(p)$ antager sit globale maksimum og indse at de to grafer giver samme resultat.

Øvelse 3.9: Sammenlign forskrifterne som du har bestemt i øvelse 3.3 og øvelse 3.5 og afgør om det er lettest at maksimere likelihoodfunktionen eller log-likelihood funktionen.

Likelihoodværdien

Hvad er likelihoodfunktionen? Udtrykket for likelihoodfunktion er identisk med sandsynlighedsfordelingen for binomialfordelingen, det vil sige

$$L_x(p) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}.$$

I likelihoodfunktionen betragtes sandsynlighedsparameteren som variablen med definitionsområde [0,1]. Observationen x skal opfattes som en konstant idet likelihoodfunktionen forudsætter en stikprøve.

Hvad er betydningen af likelihoodværdien?

Ordet likelihood kommer af det engelsk ord *"likely"*, som refererer til det der er mest sandsynligt. Hvis estimatet for sandsynlighedsparameteren er $\hat{p} = 0{,}1$ kaldes tallet $L_x(0{,}1)$ for en likelihoodværdi og angiver hvor godt estimatet understøtter stikprøven.

En høj likelihoodværdi betyder, at den observerede stikprøve er meget sandsynlig for det givne estimat. En lav likelihoodværdi betyder, at med det valgte estimat vil den observerede stikprøve sjældent forekomme.

I eksemplet med vælgertilslutningen til Socialdemokratiet, betyder en høj likelihoodværdi, at estimatet $\hat{p}$ er "godt" til at beskrive fordelingen af socialdemokratiske partivalg. Hvis likelihoodværdien er lav, vil estimatet være "dårligt" til at beskrive fordelingen af socialdemokratiske partivalg.

Bemærk, at en høj likelihoodværdi for et estimat ikke er ensbetydende med at estimatet er troværdigt i den betydning, at $\hat{p}$ er præcist lig med den sande værdi for p. Dette skyldes, at der kan være stor usikkerhed forbundet med at udtage en stikprøve.

Øvelse 3.10: Overvej hvad usikkerheden forbundet med at udtage en stikprøve kommer af?

3.4 Bestemt maksimum likelihood estimatet

Maksimum likelihood estimatet for sandsynlighedsparameteren i en binomialfordeling $\hat{p}$ er den værdi af sandsynlighedsparameteren p, der bedst understøtter den observerede stikprøve x. Det betyder, at estimatet har den største likelihoodværdi blandt alle de mulige værdier for sandsynlighedsparameteren p. Vi fortsætter eksemplet om vælgertilslutning til Socialdemokratiet.

Øvelse 3.11: Bestem likelihoodværdierne $L_{322}(0{,}1)$, $L_{322}(0{,}2)$ og $L_{322}(0{,}3)$ ved udregning.

$$\boldsymbol{L_{322}(0.1) = \binom{1635}{322} \cdot 0.1^{322} \cdot (1 - 0.1)^{1635-322} = \ldots}$$

$$\boldsymbol{L_{322}(0.2) = \ldots}$$

$$\boldsymbol{L_{322}(0.3) = \ldots}$$

Brug likelihoodværdierne til at vurdere hvilket estimat $\hat{p}_1 = 0{,}1$, $\hat{p}_2 = 0{,}2$ og $\hat{p}_3 = 0{,}3$, som bedst understøtter stikprøven med observationen $x = 322$.

Øvelse 3.12: Anvend differentialregning til at bestemme ekstremumsstedet for log-likelihood funktionen $l_{322}(\mathrm{p})$ og argumenter for, at der her er et globalt maksimum ved at lave en fortegnstabel for den afledte til log-likelihoodfunktionen (evt. med CAS-værktøj).

Øvelse 3.13: Forklar hvad maksimum likelihood estimatet fortæller om vælgertilslutningen til Socialdemokratiet i Danmark.

4. Sammenligning af to hypoteser

4.1 Hypoteser

En likelihoodfunktion kan bruges til at sammenligne to hypoteser om en population. Likelihoodfunktionen giver en likelihoodværdi for hver hypotese og fortæller dermed hvilken hypotese der bedst understøtter den observerede stikprøve. En hypotese er blot en formodning, der vedrører den egenskab om populationen, som vi undersøger. Vi vil til at begynde med formulere nogle hypoteser i ord.

Hypotese 1: *Socialdemokratiet får* 25 *procent af stemmerne til folketingsvalget.*
Hypotese 2: *Socialdemokratiet får* 30 *procent af stemmerne til folketingsvalget.*

Hypotese 1: *Gennemsnittet af fodlængden for danske kvinder er 245 mm.*
Hypotese 2: *Gennemsnittet af fodlængden for danske kvinder er 253 mm.*

I en statistisk model, kan ovenstående hypoteser angives som et estimat. Dette skyldes, at hypoteserne giver et specifikt bud på en talværdi for den ukendte parameter i en given model.
Under antagelsen om, at socialdemokratiske partivalg kan beskrives med en binomialfordeling med ukendt sandsynlighedsparameter p, kan ovenstående hypoteser oversættes til estimaterne:

Socialdemokratiet får 25 procent af stemmerne til folketingsvalget $\leftrightarrow$ $H\colon\ \hat{p}_1 = 0{,}25$,
Socialdemokratiet får 30 procent af stemmerne til folketingsvalget $\leftrightarrow$ $H\colon\ \hat{p}_2 = 0{,}3$.

Ligeså under antagelsen om at fodlængden hos danske kvinder kan beskrives ved en normalfordeling med ukendt middelværdiparameter μ, kan ovenstående hypoteser oversættes til estimaterne:

Gennemsnittet af fodlængden for danske kvinder er 245 mm $\leftrightarrow$ $H\colon\ \hat{\mu}_1 = 245$.
Gennemsnittet af fodlængden for danske kvinder er 253 mm $\leftrightarrow$ $H\colon\ \hat{\mu}_2 = 253$.

4.2 Terningekast

Du kan forestille dig, at du spiller med en terning sammen med dine venner, og at I slår mange seksere. Du får mistænke om at der er snydt med terningen, fordi den ofte slår 6. Dette skyldes, at I har slået 7 seksere ud af i alt 10 slag med terningen. Du vil gerne finde ud af hvor stor sandsynligheden er for at slå en sekser med terningen og derved undersøge din hypotese om snyd.
Dette scenarie vil vi følgende beskrive med en binomialfordeling, hvor "succes" er at slå en sekser og "fiasko" er at slå et hvilket som helst andet slag.

Statistisk model: Lad X være antallet af seksere slået ud af i alt $n = 10$ slag. X er $bin(10, p)$-fordelt, hvor p er sandsynligheden for at slå en sekser og p er ukendt.

For at undersøge hypotesen om snyd opstilles to hypoteser

- **Hypotese 1:** Vi formoder at terningen er en snydeterning med sandsynlighed $\hat{p}_1 = \frac{5}{6}$ for at slå en sekser.
- **Hypotese 2:** Vi formoder at terningen er fair med sandsynlighed $\hat{p}_2 = \frac{1}{6}$ for at slå en sekser.

Vi kan anvende likelihoodfunktionen i vores statistiske undersøgelse af terningekastene til at sammenholde de to hypoteser. For $x = 7$ og $n = 10$ bestemmer vi først likelihoodfunktionen $L_7(p)$ ud fra punktsandsynligheden for binomialfordelingen.

$$L_7(p) = \binom{10}{7} \cdot p^7 \cdot (1 - p)^3$$

Vi vil nu bestemme værdierne af likelihoodfunktionen $L_7(p)$ for de to hypoteser $\hat{p}_1 = \frac{5}{6}$ og $\hat{p}_2 = \frac{1}{6}$ ved følgende udregning.

$$L_7(\hat{p}_1) = \binom{10}{7} \cdot \left(\frac{5}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^3 = 0{,}155$$

$$L_7(\hat{p}_2) = \binom{10}{7} \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^3 = 0{,}00025$$

Det vil sige $L_7(\hat{p}_1) > L_7(\hat{p}_2)$. Dette forhold betyder, at hypotese 1 med estimatet $\hat{p}_1 = \frac{5}{6}$ understøtter vores stikprøve $x = 7$ bedre end hypotese 2.
Hovedresultatet er, at vi kan bruge likelihoodfunktionen til at undersøge hvilken af to hypoteser der stemmer bedst overens med stikprøven. Bemærk dog, at dette ikke betyder, at vi kan forkaste hypotesen med den lave likelihoodværdi.

4.3 Sammenligning af to hypoteser $\hat{\theta}_1$ og $\hat{\theta}_2$

I en generel statistisk model benævnes parameteren med θ og en hypotese vil benævnes som et estimat $\hat{\theta}$. Så snart vi har angivet sandsynlighedsfordelingen for modellen kan vi være mere konkrete. For eksempel kan vi for binomialfordelingen kalde parameteren p og hypotesen for $\hat{p}$.

I det generelle tilfælde kan to hypoteser $\hat{\theta}_1$ og $\hat{\theta}_2$ sammenlignes ved at bestemme deres likelihood-værdi og sammenholde dem.

Definition 2: Givet to hypoteser $\hat{\theta}_1$ og $\hat{\theta}_2$, hvor hypotesen $\hat{\theta}_1$ har en større likelihoodværdi end hypotese $\hat{\theta}_2$, således at

$$L_x(\hat{\theta}_1) > L_x(\hat{\theta}_2),$$

da gælder det at hypotesen $\hat{\theta}_1$ bedre understøtter stikprøven bedre end hypotesen $\hat{\theta}_2$ gør.

4.4 Likelihood ratio

Vi kan også sammenligne to hypoteser ved at bestemme *likelihood ratio*-størrelsen, som angiver forholdet mellem to likelihoodværdier og benævnes LR.

Definition 3: Givet to hypoteser $\hat{\theta}_1$ og $\hat{\theta}_2$ er likelihood ratio-størrelsen

$$LR = \frac{L_x(\hat{\theta}_1)}{L_x(\hat{\theta}_2)}.$$

Vurdering af likelihood ratio-størrelsen

Om hypotese 1, $\hat{\theta}_1$, og hypotese 2, $\hat{\theta}_2$, siges at

- Hvis likelihood ratio er tilnærmelsesvis 1, så giver hypoteserne næsten lige gode beskrivelser af den observerede stikprøve.
- Hvis likelihood ratio er mindre end 1, så giver hypotese 2 en bedre beskrivelse af den observerede stikprøve end hypotese 1.

- Hvis likelihood ratio er større end 1, så giver hypotese 1 en bedre beskrivelse af den observerede stikprøve end hypotese 2.
- Hvis likelihood ratio er tilnærmelsesvis nul, så giver hypotese 1 en meget dårlig beskrivelse af stikprøven.
- Hvis likelihood ratio er meget stor, så giver hypotese 2 en meget dårlig beskrivelse af stikprøven.

Vender vi tilbage til eksemplet om terningekastet kan likelihood ratio for de to hypoteser $\hat{p}_1 = \frac{5}{6}$ og $\hat{p}_2 = \frac{1}{6}$ bestemmes til

$$\frac{L_7(\hat{p}_1)}{L_7(\hat{p}_2)} = \frac{\binom{10}{7} \cdot \left(\frac{5}{6}\right)^7 \cdot \left(\frac{1}{6}\right)^3}{\binom{10}{7} \cdot \left(\frac{1}{6}\right)^7 \cdot \left(\frac{5}{6}\right)^3} = \frac{0{,}155}{0{,}00025} = 620.$$

Konklusion er at den observerede stikprøve $x = 7$ er hele 620 gange mere sandsynlig under hypotesen $\hat{p}_1$ end under hypotesen $\hat{p}_2$. Der er altså ret stærk statistisk evidens for hypotese 1.

4.5 Opgave om møntkast

I et binomialforsøg slås der med en mønt 10 gange, hvor udfaldet krone er succes. Mønten lander på krone i 8 af slagene og på plat i 2 af slagene. I forsøget kender vi ikke sandsynlighedsparameteren p for at slå krone. En hypotese er at mønten er symmetrisk, hvor $\hat{p}_S$ angiver estimatet for denne hypotese.

a) Angiv estimatet for hypotesen $\hat{p}_S$ for at mønten er symmetrisk.

b) Opstil en anden selvvalgt hypotese og angiv estimatet $\hat{p}_A$.

c) Bestem likelihoodfunktionen for binomialmodellen med den givne stikprøve ($x = 8$).

d) Bestem likelihoodværdien for hypotesen om at mønten er symmetrisk og for den selvvalgte hypotese.

e) Bestem likelihood ratio $\frac{L_8(\hat{p}_A)}{L_8(\hat{p}_S)}$ og forklar dens betydning i forhold til de to hypoteser.

4.6 Opgave om grafisk aflæsning af likelihoodfunktionen

En teknik der anvendes til at tælle individer i en population, går ud på at mærke en delmængde af populationen. Man udtager derefter en stikprøve og noterer hvor mange af individerne i stikprøven, som er blevet mærket tidligere. Denne metode bruges blandt andet til at tælle vilde dyr.
Andelen af vilde dyr som er henholdsvis 'mærket' og 'ikke-mærket' giver et estimat for populationens størrelse N. Metoden bruges til at tælle isfugle i Danmark, hvor man mærker 30 isfugle og sætter dem ud i naturen igen. Herefter udtages en stikprøve på 50 isfugle, hvoraf 3 isfugle er mærket og 47 isfugle ikke er mærket.

Den hypergeometriske sandsynlighedsfordeling

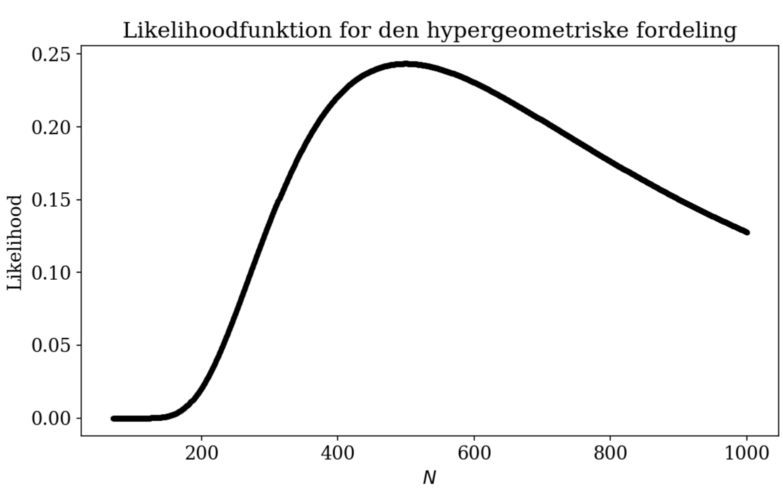
Den hypergeometriske sandsynlighedsfordeling er

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

hvor X er en stokastisk variabel der angiver antallet af succeser i en stikprøve med n individer udtrukket tilfældigt fra en population med N individer, med k succeser.

N er isfuglepopulationens størrelse i Danmark, som er ukendt. Isfugle med mærke er $k = 30$. Stikprøvestørrelsen er $n = 50$ og i stikprøven var $x = 3$ isfugle mærket.

Vi antager, at isfuglene er udvalgt/talt tilfældigt, og da kan likelihoodfunktionen for N bestemmes ved hjælp af den hypergeometriske sandsynlighedsfordeling $L(N) = P(X = 3) = \frac{\binom{30}{3}\binom{N-30}{47}}{\binom{N}{50}}$.
Nedenfor ses likelihood funktionen af parameteren N.



Figur 5: Likelihoodfunktion for den hypergeometrisk sandsynlighedsfordeling.

a) Bestem maksimum likelihood estimatet for N ved at aflæse på grafen for likelihood funktionen.

b) Giv en betydning af det aflæste estimat.

c) Vurdér de fire bud på estimater $\widehat{N}_1 = 200$, $\widehat{N}_2 = 350$, $\widehat{N}_3 = 400$ og $\widehat{N}_4 = 600$ ved at aflæse på grafen (så nøjagtigt som muligt).

d) Hvilke betingelser bør være opfyldt om de mærkede fugle i populationen for, at estimatet er troværdigt?

5. Betydning af stikprøvestørrelsen n

5.1 Hvad sker der når stikprøvestørrelsen n vokser?

I en statistisk undersøgelse har stikprøvestørrelsen n betydning for præcisionen og troværdigheden af et estimat $\hat{\theta}$ i en statistisk model. Har vi en stor stikprøve, kan vi som regel komme med meget troværdige og sikre konklusioner om populationen. Har vi derimod en lille stikprøve bliver vores konklusioner usikre og vi kan måske slet ikke konkludere noget som helst om populationen på denne baggrund.

Vi vender tilbage til opgaven om møntkast, hvor vi jo kun slog 10 gange med mønten. Det kan derfor være svært at vide om det blot var tilfældigt, at vi slog krone hele 8 gange. Vi vil nok stadig være tilbøjelige til at tro at mønten er symmetrisk. Vi udvider forsøget og slår $n = 50$ gange med mønten i stedet for, slår krone 40 af gangene ($x = 40$) og opstiller en ny statistisk model: X er antal kroner i 50 kast og $X \sim bin(50, p)$. Vi skal nu undersøge hvad der sker, når stikprøvestørrelsen bliver større.

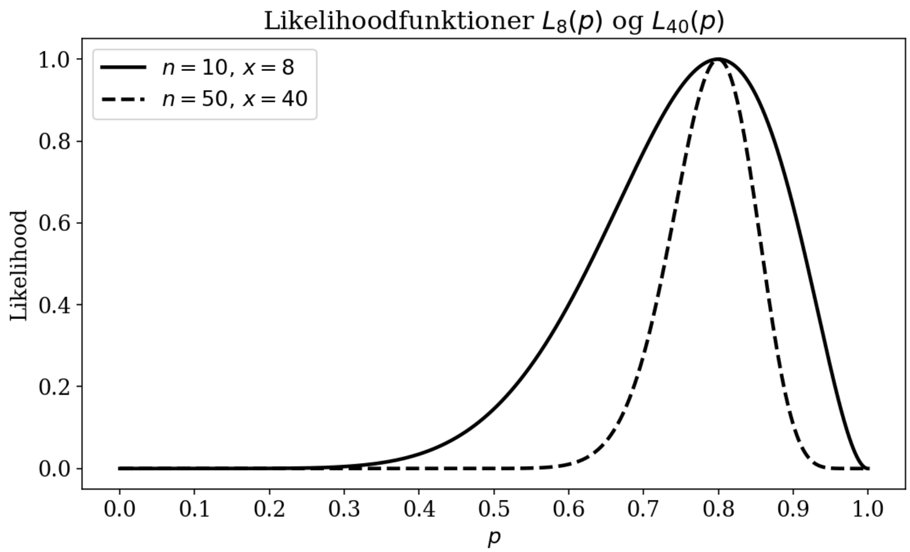
5.2 Den grafiske betydning af stikprøvestørrelsen n

Betydningen af stikprøvestørrelsen/antalsparameteren n har betydning for grafen for likelihood-funktionen. Jo større stikprøvestørrelse n er, jo smallere og stejlere bliver grafen for likelihoodfunktionen omkring maksimum likelihood estimatet.

Nedenunder ses graferne for likelihoodfunktionerne $L_8(p) = \binom{10}{8} \cdot p^8 \cdot (1-p)^2$ og

$L_{40}(p) = \binom{50}{40} \cdot p^{40} \cdot (1-p)^{10}$. Af graferne for likelihoodfunktionen kan man aflæse likelihood-værdien for forskellige værdier af p. På plottet ses det, at funktionerne har samme globale maksimumsværdi, hvilket giver anledning til det samme maksimum likelihood estimatet

$$\hat{p} = \frac{x}{n} = \frac{40}{50} = \frac{8}{10} = 0{,}8.$$



Figur 6: Likelihoodfunktionerne $L_8(p)$ og $L_{40}(p)$

For en lille stikprøve (med $n = 10$ og $x = 8$) vil grafen som regel være bred og flad. Det betyder, at de forskellige mulige værdier for p har en nogenlunde ens likelihoodværdi.

Maksimum likelihood estimatet er som sagt den værdi for p som understøtter stikprøven bedst.
Når de andre mulige værdier for parameteren p næsten er lige så gode og har likelihoodværdier tæt på hinanden, siger vi at den statistiske evidens for maksimum likelihood estimatet er svagt og LR er tæt på 1.

For en større stikprøve ($n = 50$ og $x = 40$) vil grafen for likelihoodfunktionen blive smallere. Det betyder, at likelihoodværdien formindskes markant når estimatet $\hat{p}$ bevæger sig væk fra det globale maksimum og dermed vil LR bliver markant forskellig fra 1. Altså vil der være en store fordel ved at vælge maksimum likelihood estimatet.

5.3 Stikprøvestørrelsens effekt på likelihood ratio

Effekten af stikprøvestørrelsen n påvirker likelihood ratio-størrelsen. I mønteksemplet kan vi sammenligne de to hypoteser: $\hat{p}_1 = 0{,}75$ (der er 0,75 sandsynlighed for at slå krone) og $\hat{p}_2 = 0{,}5$ (mønten er symmetrisk).
Når stikprøven er lille ($n = 10$ og $x = 8$) bestemmes likelihood ratio til

$$LR = \frac{L_8(\hat{p}_1)}{L_8(\hat{p}_2)} = \frac{\binom{10}{8} \cdot (0{,}75)^8 \cdot (0{,}25)^2}{\binom{10}{8} \cdot (0{,}5)^8 \cdot (0{,}5)^2} = 6{,}41.$$

Den observerede stikprøve $x = 8$ er hermed $6{,}41$ gange mere sandsynlig under hypotesen $\hat{p}_1 = 0{,}75$ end under hypotesen $\hat{p}_2 = 0{,}5$.

Når stikprøven er stor ($n = 50$ og $x = 40$) bestemmes likelihood ratio til

$$LR = \frac{L_{40}(\hat{p}_1)}{L_{40}(\hat{p}_2)} = \frac{\binom{50}{40} \cdot (0{,}75)^{80} \cdot (0{,}25)^{20}}{\binom{50}{40} \cdot (0{,}5)^{80} \cdot (0{,}5)^{20}} = 10798{,}18$$

Den observerede stikprøve $x = 40$ er $10798{,}18$ gange mere sandsynlig under hypotesen $\hat{p}_1 = 0{,}75$ end under hypotesen $\hat{p}_2 = 0{,}5$. Den statistiske evidens for hypotesen $\hat{p}_1 = 0{,}75$ er dermed blevet markant større.

Øvelse 5.1: Der slås krone fem gange ($x = 5$) ved n møntkast. Bestem hvor stor n skal være for at stikprøven er lige sandsynlig under hypotesen $\hat{p}_1 = 0{,}75$ som under hypotesen $\hat{p}_2 = 0{,}5$.

5.4 Opgave om Earl og saunaen

En mand Earl er rejst på ferie til et fremmed land, hvor han bor på et hotel, som har to saunaer. En sauna kun for kvinder og en kønsneutral sauna. Earl beslutter sig en eftermiddag for at tage i sauna, men et problem opstår, da han ikke forstår det lokale sprog. Han kan ikke finde ud af, hvilken af saunaerne, som er den kønsneutrale. Forestil dig, at Earl står ude foran saunaerne og prøver at tyde skiltene på det lokale sprog, da tre kvinder uafhængigt af hinanden kommer ud af saunaen til højre.

a) Opstil en binomialmodel med udgangspunkt i Earls oplevelse med saunaerne, hvor ”succes” er antallet af kvinder som går ud af saunaen til højre og stikprøvestørrelsen er $n = 3$.

b) Earls undren om hvilken sauna han skal vælge, giver anledning til to hypoteser enten er saunaen til højre den sauna, som kun er forbeholdt kvinder $\hat{p}_K$ eller også er det den kønsneutrale sauna $\hat{p}_N = 0{,}5$. Angiv estimatet $\hat{p}_K$.

c) Hvis døren til højre er saunaen kun forbeholdt kvinder hvad er så sandsynligheden for, at alle tre personer der forlader saunaen, er kvinder. Det vil sige, at du skal bestemme $L_3(\hat{p}_K)$.

d) Hvis døren til højre er den kønsneutrale sauna hvad er så sandsynliggeden for, at alle tre personen der forlader saunaen, er kvinder. Det vil sige, at du skal bestemme $L_3(\hat{p}_N)$.

e) Bestem likelihood ration $\frac{L_3(\hat{p}_K)}{L_3(\hat{p}_N)}$ og forklar hvad resultatet fortæller om de to hypoteser $\hat{p}_K$ og $\hat{p}_N$.

f) Bestem hvor mange kvinder der skal forlade saunaen til højre før stikprøven, er 16 gange mere sandsynlig til at forekomme under hypotesen $\hat{p}_K$ end under hypotesen $\hat{p}_N$

B Note on Figures and Plots

Several figures and plots are made by me in Python or Latex as part of this thesis. A single figure has been made by hand. Some figures and plots are used to highlight subject-didactic perspectives, other plots are made for the textbook chapter and a few figures have been developed to create an overview of my didactical analyses. It has been agreed with my supervisor Carl Winsløw, that any code for creating plots or figures should not be attached to the thesis.

C Brainstorm on Accessibility

In the initial part of my work on this thesis, I conducted a brainstorm based on Kirsch's notion of "making accessible". The brainstorm can be found in table 3 below.

Intuition	**Application**
to use one's instinct	to incorporate the real world
to use physical experiences	to work with problem-solving
to use common sense	to manipulate a concept
to guess	to combine subjects
Recognizing	**Modes of representation**
to recognize patterns	to provide two explanations
to activate prior knowledge	to switch approaches
to distinguish from	to see different examples
to be aware of the next step	to be verbal and graphical

Table 3: Concept-board on 'accessibility' inspired by Kirsch .

D The Method of Mark and Recapture

To find the maximum of $L(N)$, one can determine when the fractions $\frac{L(N|r)}{L(N-1|r)}$ is strictly smaller than 1 and when it is strictly larger than 1. By the calculations:

$$\frac{L(N|r)}{L(N-1|r)} = \frac{\frac{\binom{t}{r}\binom{N-t}{k-r}}{\binom{N}{k}}}{\frac{\binom{t}{r}\binom{N-t-1}{k-r}}{\binom{N-1}{k}}} = \frac{\binom{N-t}{k-r}\binom{N-1}{k}}{\binom{N-t-1}{k-r}\binom{N}{k}} = \frac{\frac{(N-t)!}{(K-r)!(N-t-k+r)!}\frac{(N-1)!}{k!(N-k-1)!}}{\frac{(N-t-1)!}{(k-r)!(N-t-k+r-1)!}\frac{(N-1)!}{k!(N-k)!}}$$

Then by reducing the expression of the ratio, we get

$$\frac{L(N|r)}{L(N-1|r)} = \frac{(N-t)!(N-1)!(N-t-k+r-1)!(N-k)!}{(N-t-1)!N!(N.t-k+r)!(N-k-1)!} = \frac{(N-t)(N-k)}{N(N-t-k+r)}.$$

$$= \frac{(N-t)(N-k)}{N(N-t-k+r)} = \frac{N^2-Nt-Nk+tk}{N^2-Nt-Nk+Nr} = \frac{tk}{Nr}.$$

Then $\frac{tk}{Nr} > 1$ implies $\frac{tk}{r} > N$. One can see that

$$\frac{L(N)}{L(N-1)} > 1 \quad \Leftrightarrow \quad N < \frac{tk}{r}.$$

and

$$\frac{L(N)}{L(N-1)} \leq 1 \quad \Leftrightarrow \quad N \geq \frac{tk}{r}.$$

This shows that the function $L(N)$ is maximized at $N = \lfloor \frac{tk}{r} \rfloor$. So the MLE is $\hat{N}_{MLE} = \lfloor \frac{tk}{r} \rfloor$.