BACHELOR

Convergence speed of maximum likelihood estimator for boundary detection

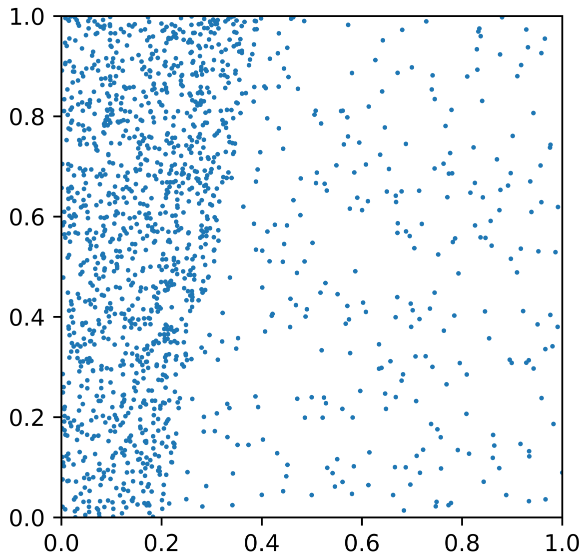
Beerens, Lucas

Award date:
2021

Department of Mathematics and Computer Science

Convergence speed of maximum likelihood estimator for boundary detection

Bachelor final project Applied Mathematics



Author:
Lucas Beerens

Supervisor:
dr. Pim van der Hoorn

June 13, 2021

Abstract

In a super-resolution imaging technique called iPAINT, interfaces between tiny oil droplets and surrounding water can be detected. The technique creates a dataset of points that lie in both the droplets and the surrounding water. There is a difference between the density of the points inside and outside of the droplets. To analyse the interfaces, we will zoom far enough into the interface such that it resembles a straight line. We model this as each point being on either side with some probability. The location on that side is uniformly distributed. We assume that the boundary is a straight line from the top to the bottom edge of a unit square. We investigate how quickly the maximum likelihood estimator for the boundary converges as the amount of points increases. We give a method to prove that it converges at least as fast as $1/n$, where n is the amount of points. This proof is based on the proof of a one-dimensional equivalent to this model. This proof is described roughly in a paper by Chernoff and Rubin. We work out the entire proof of the one-dimensional variant and give all relevant details. In the one-dimensional case, we consider the unit interval. There is a boundary, which is given by a fixed point on this interval. Points are generated that lie on either side of the boundary with some given probability. The location on the side of the boundary is uniformly distributed.

Contents

Chapter 1

Introduction

1.1 Motivation

In chemistry there is a type of mixture of substances called colloid. This is a mixture where a substance is scattered throughout a liquid in the form of small particles. These particles can essentially be seen as tiny droplets of the substance with a diameter which is of nanometre scale. An example of a colloid is milk. In milk, the particles are small beads of fat. These particles are suspended in a water-based solution. These colloids are found in many common products.

A colloid is called stable if the attraction force between particles is smaller than some threshold. If this is not the case, the particles will clump together and the mixture will cease to be a colloid. Hence in the production process of a colloid, the stability is crucial. There are many methods to improve stability. To investigate some of these methods, it is useful to visualise the interface between the particle and the surrounding liquid.

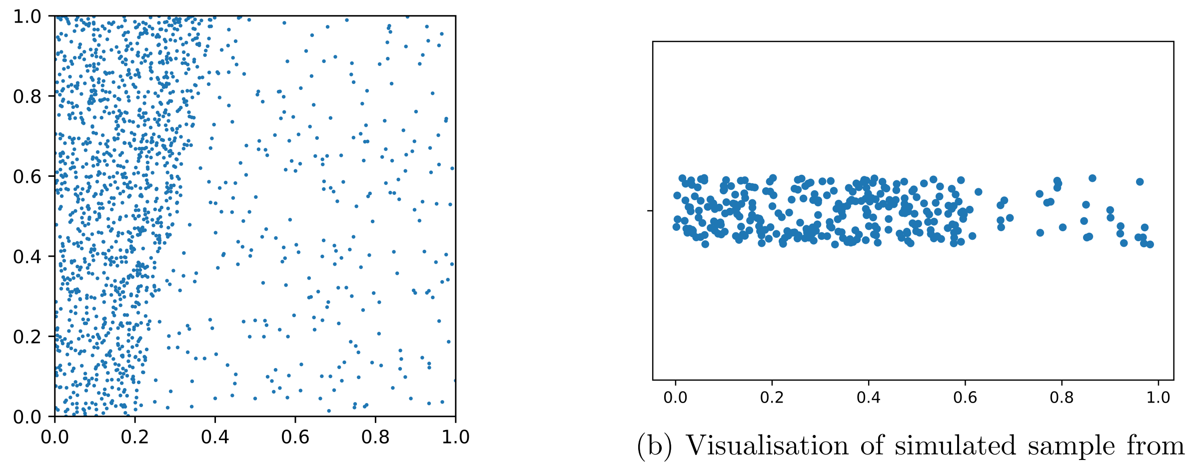
The visualisation of the interface comes with two problems: the particles are very sensitive and the scale is very small. Because of these, conventional microscopy can not be used. A method that is able to do this is iPAINT[1]. The result of applying this method is a collection of three-dimensional points. The density of the points within the particles is different from the density of the points in the surrounding liquid. In the analysis of this data, one of the important factors is the estimation of the location of the interfaces.

In the master's thesis by D. van der Haven[2], a method is described to automatically estimate the location of the interface. In this thesis, there is a focus on a two-dimensional equivalent of this problem with the observation that it can easily be generalised to a three-dimensional setting. One of the tools that are used in this paper is the maximum likelihood estimator (MLE).

The model used for the maximum likelihood estimation is as follows. When zooming in far enough on an interface, the boundary will approximately be a straight line. Hence it is assumed that the boundary is a straight line. The observations are assumed to come from an inhomogeneous Poisson point process where the density on both sides of the boundary are homogeneous, but different from each other.

1.2 Goal

One of the properties of an estimator that is useful to know is its convergence speed. In this case we will investigate how quickly the error of the estimator of the boundary decreases as the amount of observations increases. This is the topic of this thesis. However, we will not consider the exact model that was just described. We will simplify two aspects. The first simplification is that we do not assume the observations to come from an inhomogeneous Poisson. Instead we assume that they are a sample from a distribution which has different but constant densities on both sides of the line. The second simplification is that it will be that it is assumed that the boundary touches the upper and bottom edge of the image. These two assumptions allow us to look at an easier situation, while making sure that the result can likely be generalised to the model that was described before. We assume we work with the unit square. An example of a sample from this distribution can be seen in Figure 1.1a.



(a) Visualisation of simulated sample from the two-dimensional random variables

(b) Visualisation of simulated sample from the one-dimensional random variables using a strip plot

Figure 1.1: Visualisations of the two models

A one-dimensional equivalent of this problem has already been considered before by Chernoff and Rubin[3]. In this paper, the unit interval is considered with a boundary splitting it into two parts. The observations come are distribution which has different but constant densities on both sides of the boundary point. We assume we work with the unit interval. A strip plot of an example sample from this distribution can be seen in Figure 1.1b. The MLE is found to have convergence speed of $1/n$, where n is the amount of observations. This is considered to be relatively fast.

We hypothesise that the MLE in the two-dimensional case converges with the same speed. To find the convergence speed of an estimator, we need to show that something is both the upper and lower bound for the speed. In mathematical terms, if $\hat{\alpha}_n$ is the MLE for the boundary with true value α_0, then it needs to be shown that $\hat{\alpha}_n - \alpha_0 = O_p(1/n)$, but $\hat{\alpha}_n - \alpha_0 \neq o_p(1/n)$ as $n \to \infty$. The definitions of O_p and o_p are critical to the thesis and can be found in Appendix Section B, where some assumed knowledge is explained. More specifically, they are mentioned in Definitions B.6 and B.7 respectively. In this thesis we will only concern ourselves with the former statement, which says that MLE converges at least as fast as $1/n$. The goal is to describe a method to prove that this is the case in the two-dimensional case.

1.3 Approach

In order to find out how it can be proven that the error of the MLE is $O_p(1/n)$ as $n \to \infty$, we will start by looking at the one-dimensional case as described by Chernoff and Rubin[3]. The paper contains the main ideas of the method on how to prove that the error of the MLE is $O_p(1/n)$ as $n \to \infty$ in the one-dimensional case. However, most details are absent. This makes it difficult to translate the proof into the two-dimensional setting. For this reason, we start by completely working out the proof in the one-dimensional setting. Then we use this to find a method to prove the two-dimensional problem. Part of this proof is given in the appendix. The appendix also includes some prerequisite knowledge.

The outline of the proof of the one-dimensional problem is as follows. First we define all the necessary notation, before finding the MLEs for all parameters. Then it will be shown that the MLEs are consistent. In order to prove that the error of the MLE for the boundary is $O_p(1/n)$ as $n \to \infty$, another estimator shall be introduced. This estimator on its own does not have any practical use as it uses information of the true values of the parameters. However, it can be shown that the error of this new estimator is $O_p(1/n)$ as $n \to \infty$. Finally it will be shown that the difference between the new estimator and the MLE is $o_p(1/n)$ as $n \to \infty$, from which the desired result immediately follows.

Chapter 2

Convergence speed in 1D model

In this chapter, we will consider the convergence speed of the one-dimensional model. The random variable that is considered in this model has different constant densities in two parts of the unit interval. We split the unit interval into two sections based on a parameter $\alpha \in (0, 1)$. This gives us the intervals $[0, \alpha]$ and $(\alpha, 1]$. The random variable lies in these intervals with probabilities θ and $1 - \theta$ respectively. Within these intervals, the random variable is uniformly distributed. This gives us a probability distribution function of the form

$$f(x) = \begin{cases} \beta, & \text{if } x \in [0, \alpha], \\ \gamma, & \text{if } x \in (\alpha, 1], \\ 0, & \text{otherwise,} \end{cases} \tag{2.0.1}$$

where

$$\int_0^\alpha f(x)dx = \alpha\beta = \theta, \tag{2.0.2}$$

and

$$\int_\alpha^1 f(x)dx = (1 - \alpha)\gamma = 1 - \theta. \tag{2.0.3}$$

This naturally leads us to define β and γ as functions of α and θ. The functions $\beta, \gamma : (0, 1)^2 \to (0, \infty)$ are defined by

$$\beta(\alpha, \theta) := \frac{\theta}{\alpha}, \quad \text{and} \quad \gamma(\alpha, \theta) := \frac{1 - \theta}{1 - \alpha}. \tag{2.0.4}$$

Now these are defined such that for all $\alpha, \theta \in (0, 1)$ replacing β and γ in Equation (2.0.1) with their respective functions from Equation (2.0.4) yields a proper probability density function. For fixed parameters $\alpha, \theta \in (0, 1)$ the probability density function $f : \mathbb{R} \to [0, \infty)$ is defined by

$$f(x, \alpha, \theta) := \begin{cases} \beta(\alpha, \theta), & \text{if } x \in [0, \alpha], \\ \gamma(\alpha, \theta), & \text{if } x \in (\alpha, 1], \\ 0, & \text{otherwise.} \end{cases} \tag{2.0.5}$$

We are interested in maximum likelihood estimation. Therefore we need to sample from the distribution. To this end we fix $\alpha_0, \theta_0 \in (0,1)$ and let $X_1, X_2, \dots$ be i.i.d. distributed with density $f(x, \alpha_0, \theta_0)$. These samples will be used by the estimator.

Assumption 2.1. *Since it is assumed that f is not the PDF for the uniform distribution and hence the densities on the two sides of the boundary are different, we can also assume that $\alpha_0 \neq \theta_0$.*

Since we will often work with the density $f(x, \alpha_0, \theta_0)$ of the random variables $(X_i)_{i \in \mathbb{N}}$ we define the values $\beta_0, \gamma_0 \in \mathbb{R}^+$ as

$$\beta_0 := \beta(\alpha_0, \theta_0), \qquad \text{and} \qquad \gamma_0 := \gamma(\alpha_0, \theta_0). \tag{2.0.6}$$

Assumption 2.2. *Without loss of generality we assume that $\beta_0 > \gamma_0$.*

We now want to find the CDF corresponding to the distribution with PDF $f(x, \alpha_0, \theta_0)$. Suppose X is a random variable with PDF $f(x, \alpha_0, \theta_0)$. Then for $x \in [0, \alpha_0]$ we have

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t, \alpha_0, \beta_0) dt \tag{2.0.7}$$

$$= \int_{0}^{x} \beta_0 dt \tag{2.0.8}$$

$$= \beta_0 \cdot x, \tag{2.0.9}$$

and for $x \in (\alpha_0, 1]$ we have

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t, \alpha_0, \beta_0) dt \tag{2.0.10}$$

$$= 1 - \int_{x}^{1} \gamma_0 dt \tag{2.0.11}$$

$$= 1 - \gamma_0(1 - x). \tag{2.0.12}$$

Therefore the CDF corresponding to $f(x, \alpha_0, \theta_0)$ is

$$F_0(x) := \begin{cases} 0, & \text{if } x < 0, \\ \beta_0 x, & \text{if } x \in [0, \alpha_0], \\ 1 - \gamma_0(1 - x), & \text{if } x \in (\alpha_0, 1], \\ 1, & \text{if } x > 1. \end{cases} \tag{2.0.13}$$

It will also be necessary to look at the empirical cumulative distribution function. For any $n \in \mathbb{N}$ the eCDF for $X_1, \dots, X_n$ is denoted by

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \leq x}. \tag{2.0.14}$$

2.1 Maximum likelihood estimators

The goals of this section are to find an expression for the MLE of θ_0 in terms of the MLE of α_0 and to find a convenient form for the MLE of α_0. To compute the MLEs, we first look at the likelihood

$$\mathcal{L}_n(\alpha,\theta) = \prod_{i=1}^{n} f(X_i,\alpha,\theta) \tag{2.1.1}$$

$$= \prod_{X_i<\alpha} \beta(\alpha,\theta) \prod_{X_i>\alpha} \gamma(\alpha,\theta) \tag{2.1.2}$$

$$= (\beta(\alpha,\theta))^{nF_n(\alpha)}(\gamma(\alpha,\theta))^{n(1-F_n(\alpha))}. \tag{2.1.3}$$

Therefore the log-likelihood divided by n is given by

$$\ell_n(\alpha,\theta) = \log\left(\frac{\theta}{\alpha}\right) F_n(\alpha) + \log\left(\frac{1-\theta}{1-\alpha}\right)(1-F_n(\alpha)). \tag{2.1.4}$$

First we want to find the MLE for θ. The partial derivative of Equation (2.1.4) with respect to θ is given by

$$\frac{\partial \ell_n}{\partial \theta}(\alpha,\theta) = \frac{F_n(\alpha)}{\theta} - \frac{1-F_n(\alpha)}{1-\theta}. \tag{2.1.5}$$

The zero of this equation is given by $F_n(\alpha)$. To check whether this is actually the MLE, we compute the second order partial derivative of Equation (2.1.4) with respect to θ and verify whether it is negative. The second order partial derivative is

$$\frac{\partial^2 \ell_n}{\partial \theta^2}(\alpha,\theta) = -\frac{F_n(\alpha)}{\theta^2} - \frac{1-F_n(\alpha)}{(1-\theta)^2}. \tag{2.1.6}$$

Evaluation in $\theta = F_n(\alpha)$ yields

$$\frac{\partial^2 \ell_n}{\partial \theta^2}(\alpha,F_n(\alpha)) = -\frac{1}{F_n(\alpha)} - \frac{1}{1-F_n(\alpha)} \tag{2.1.7}$$

$$= \frac{-1}{F_n(\alpha)(1-F_n(\alpha))}. \tag{2.1.8}$$

We know that $F_n(\alpha)(1-F_n(\alpha)) > 0$ for any value of α and therefore Equation (2.1.8) is negative. We can conclude that

$$\hat{\theta}_n := F_n(\hat{\alpha}_n) \tag{2.1.9}$$

is the MLE for θ_0, where $\hat{\alpha}_n$ is the MLE for α_0.

Now we want to find the MLE for α_0. To do this, we will fill in the MLE for θ_0 into Equation (2.1.4). This yields

$$\ell_n(\alpha,F_n(\alpha)) = \log\left(\frac{F_n(\alpha)}{\alpha}\right) F_n(\alpha) + \log\left(\frac{1-F_n(\alpha)}{1-\alpha}\right)(1-F_n(\alpha)). \tag{2.1.10}$$

Since the MLE maximises this function, we can write

$$\hat{\alpha}_n := \underset{x\in(0,1)}{\arg\max} \left(\log\left(\frac{F_n(x)}{x}\right) F_n(x) + \log\left(\frac{1-F_n(x)}{1-x}\right)(1-F_n(x)) \right). \tag{2.1.11}$$

This is the expression for $\hat{\alpha}_n$ that we shall use in the next section to prove consistency.

2.2 Consistency of MLEs

The most essential property of an estimator is consistency. Not all MLEs are consistent though. As this property is necessary for the rest of this chapter, we will show that both $\hat{\alpha}_n$ and $\hat{\theta}_n$ are consistent.

Theorem 2.1. *The MLE $\hat{\theta}_n = F_n(\hat{\alpha}_n)$, which estimates θ_0, is a consistent estimator for θ_0.*

Theorem 2.2. *The MLE $\hat{\alpha}_n$, which estimates α_0, is consistent.*

Theorem 2.1 will follow straightforwardly from Theorem 2.2 and the Glivenko-Cantelli theorem, which can be found in the appendix at Theorem B.4. Since the this proof is relatively short in comparison to the proof of Theorem 2.2, we start with the proof of Theorem 2.1.

2.2.1 Proof of Theorem 2.1

Proof of Theorem 2.1. This proof shall make a straightforward use of the triangle inequality, combined with the Glivenko-Cantelli theorem and consistency of $\hat{\alpha}_n$.

Let $\epsilon_1, \epsilon_2 > 0$. Since

$$F_0(\alpha_0) = \beta_0\alpha_0 = \theta_0, \tag{2.2.1}$$

we find that

$$\begin{aligned} \mathbb{P}(|F_n(\hat{\alpha}) - \theta_0| < \epsilon_1) &= \mathbb{P}(|F_n(\hat{\alpha}) - F_0(\alpha_0)| < \epsilon_1) &(2.2.2)\\ &\geq \mathbb{P}(|F_n(\hat{\alpha}) - F_0(\hat{\alpha})| + |F_0(\hat{\alpha}) - F_0(\alpha_0)| < \epsilon_1). &(2.2.3)\end{aligned}$$

By the Glivenko-Cantelli theorem, which is stated in Theorem B.4, we have

$$\sup_{x\in(0,1)} \left|F_n(x) - F_0(x)\right| \xrightarrow{\text{a.s.}} 0. \tag{2.2.4}$$

Since almost sure convergence implies convergence in probability, we can choose $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{x\in(0,1)} \left|F_n(x) - F_0(x)\right| < \frac{\epsilon_1}{2}\right) > 1 - \frac{\epsilon_2}{2}. \tag{2.2.5}$$

Since

$$\sup_{x\in(0,1)} \big|F_n(x) - F_0(x)\big| < \frac{\epsilon_1}{2} \implies |F_n(\hat{\alpha}) - F_0(\hat{\alpha})| < \frac{\epsilon_1}{2}, \tag{2.2.6}$$

we find that for all $n \geq N_1$ we have

$$\mathbb{P}\left(\big|F_n(\hat{\alpha}) - F_0(\hat{\alpha})\big| < \frac{\epsilon_1}{2}\right) > \mathbb{P}\left(\sup_{x\in(0,1)} \big|F_n(x) - F_0(x)\big| < \frac{\epsilon_1}{2}\right) > 1 - \frac{\epsilon_2}{2}. \tag{2.2.7}$$

By convergence of $\hat{\alpha}$ to α_0 in probability for $n \to \infty$ and continuity of F_0 we can apply the continuous mapping theorem, which is stated in Theorem B.3, to find that

$$F_0(\hat{\alpha}) = F_0(\alpha_0) + o_p(1). \tag{2.2.8}$$

Hence we can choose $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$ we have

$$\mathbb{P}\left(|F_0(\hat{\alpha}) - F_0(\alpha_0)| < \frac{\epsilon_1}{2}\right) > 1 - \frac{\epsilon_2}{2}. \tag{2.2.9}$$

If

$$|F_0(\hat{\alpha}) - F_0(\alpha_0)| < \frac{\epsilon_1}{2} \quad \text{and} \quad \big|F_n(x) - F_0(x)\big| < \frac{\epsilon_1}{2}, \tag{2.2.10}$$

then

$$|F_n(\hat{\alpha}) - F_0(\hat{\alpha})| + |F_0(\hat{\alpha}) - F_0(\alpha_0)| < \epsilon_1. \tag{2.2.11}$$

Hence by the Fréchet inequalities from Lemma B.6 we find that for $n \geq \max\{N_1, N_2\}$

$$\begin{aligned}
\mathbb{P}(|F_n(\hat{\alpha}) - \theta_0| < \epsilon_1) &\geq \mathbb{P}(|F_n(\hat{\alpha}) - F_0(\hat{\alpha})| + |F_0(\hat{\alpha}) - F_0(\alpha_0)| < \epsilon_1) && (2.2.12)\\
&\geq \mathbb{P}\left(|F_0(\hat{\alpha}) - F_0(\alpha_0)| < \frac{\epsilon_1}{2}, \quad \big|F_n(x) - F_0(x)\big| < \frac{\epsilon_1}{2}\right) && (2.2.13)\\
&\geq 1 - \epsilon_2, && (2.2.14)
\end{aligned}$$

which concludes the proof of the theorem. ■

2.2.2 Proof of Theorem 2.2

In order to prove Theorem 2.2, which states that $\hat{\alpha}_n$ is a consistent estimator for α_0, we first introduce some relevant notation. Then we will explain the proof heuristically while stating the necessary lemmas. After this the details for the proof of the theorem will be given.

For any function $G : \mathbb{R} \to [0, 1]$ we define

$$\Phi(G) = \operatorname*{arg\,max}_{x\in(0,1)} \left(G(x) \log\left(\frac{G(x)}{x}\right) + (1 - G(x)) \log\left(\frac{1 - G(x)}{1 - x}\right)\right). \tag{2.2.15}$$

Using this notation, we can write

$$\hat{\alpha}_n = \Phi(F_n). \tag{2.2.16}$$

We will use the following lemma.

Lemma 2.3. *Let $G_n(x)$ be a sequence of nonrandom c.d.f.'s such that*

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = O(1), \tag{2.2.17}$$

$$\sup_{x\in(0,1)} \left|\frac{1-G_n(x)}{1-x}\right| = O(1), \tag{2.2.18}$$

and

$$\sup_{x\in(0,1)} |G_n(x) - F_0(x)| = o(1), \tag{2.2.19}$$

where o and O are for $n \to \infty$. Then

$$\Phi(G_n) = \alpha_0 + o(1), \tag{2.2.20}$$

for $n \to \infty$.

Lemma 2.3 will be proven after we finish proving Theorem 2.2. We use this lemma to get to the next lemma. Lemma 2.3 only mentions deterministic CDFs. In our case, we are talking about stochastic CDFs F_n. Fortunately, it is possible to directly translate the lemma into random CDFs. This is the case because the calculus of o and O is the same as the calculus of o_p and O_p. A more precise statement can be found in Theorem 1 of [4]. The proof can be found in Corollary 1 of [5]. This gives us a new lemma that we will state without proof, as it is a direct result from the equivalent calculus.

Lemma 2.4. *Let $G_n(x)$ be a sequence of random c.d.f.'s such that*

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = O_p(1), \tag{2.2.21}$$

$$\sup_{x\in(0,1)} \left|\frac{1-G_n(x)}{1-x}\right| = O_p(1), \tag{2.2.22}$$

and

$$\sup_{x\in(0,1)} |G_n(x) - F_0(x)| = o_p(1). \tag{2.2.23}$$

Then

$$\Phi(G_n) = \alpha_0 + o_p(1). \tag{2.2.24}$$

Applying Lemma 2.4 to the sequence F_n would lead to the desired result of Theorem 2.2. Therefore, we shall check whether F_n satisfies the conditions in Equations (2.2.21), (2.2.22), and (2.2.23).

The third condition is a direct consequence from the Glivenko-Cantelli theorem as stated in Theorem B.4. We now consider the first condition. We want to be able to consider a uniform distribution instead of F_n. To do this we first notice that

$$\sup_{x\in(0,1)} \left|\frac{F_n(x)}{x}\right| = \sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| + \sup_{x\in(\alpha_0,1)} \left|\frac{F_n(x)}{x}\right|. \tag{2.2.25}$$

The supremum over $(\alpha_0, 1)$ is bounded $1/\alpha_0$. Thus it suffices to show that

$$\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| = O_p(1), \tag{2.2.26}$$

as $n \to \infty$. To do this we introduce a new random variable Y_i that is equal to $\beta_0 X_i$ if $X_i \leq \alpha_0$ and equal to some uniformly distributed random variable on $[\theta_0, 1]$ otherwise. Then Y_i is uniformly distributed on the unit interval. However we also have

$$\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| \leq \beta_0 \sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right|, \tag{2.2.27}$$

where G_n is the eCDF for Y_i. When can then apply the following lemma, which is proven in Appendix Section A.1. The proof is not included here as it is not directly relevant for the generalisation to two dimensions.

Lemma 2.5. *If G_n is the eCDF for a random variable which is uniformly distributed on $[0, 1]$, then*

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = O_p(1), \tag{2.2.28}$$

for $n \to \infty$.

We then conclude that Equation (2.2.26) holds and therefore the first condition is met. The second condition follows by using that the first condition holds regardless of the parameters α_0 and θ_0. We introduce $Z_i = 1 - X_i$ and notice that

$$\sup_{x\in(0,1)} \frac{1 - F_n(x)}{1 - x} = \sup_{x\in(0,1)} \frac{1 - F_n(1 - x)}{1 - (1 - x)}. \tag{2.2.29}$$

Since Z_i is in the same family of random variables as X_i, the second condition also holds. We will now give the detailed proof.

Proof of Theorem 2.2. Due to Lemma 2.4, we only need to show that F_n satisfies the conditions in Equations (2.2.21), (2.2.22), and (2.2.23).

We start with the third condition, as it can be verified very quickly. It is known that

$$\sup_{x\in(0,1)} \left|F_n(x) - F_0(x)\right|, \tag{2.2.30}$$

converges to 0 almost surely by the Glivenko-Cantelli theorem, which is stated in Theorem B.4. However, almost sure convergence implies convergence in probability. Hence the third condition is met.

Now we will look at the first condition. The first step will be to show that we can actually consider the eCDF of a uniform distribution instead of F_n. By some technical lemma that will be proven after this theorem, the condition will hold. For the second condition, we can use the result for the first condition by rewriting the problem.

We start by showing that we can consider uniform distributions. To do this we will show that we only need to care about

$$\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right|, \tag{2.2.31}$$

instead of

$$\sup_{x\in(0,1)} \left|\frac{F_n(x)}{x}\right|. \tag{2.2.32}$$

Then we will define a new random variable which is uniformly distributed on $[0,1]$ such that its eCDF is equal to $F_n(x/\beta_0)$ for $x \leq \alpha_0$. The supremum over $(0,\alpha_0)$ of $\left|\frac{F_n(x)}{x}\right|$ will then be bounded from above by some scalar multiple of the supremum over $[0,1]$ of the eCDF for the uniform distribution. Lemma 2.5 will then show that the condition is met.

For any $M > 0$ we can use the Fréchet inequalities from Lemma B.6 to find that

$$\mathbb{P}\left(\sup_{x\in(0,1)} \left|\frac{F_n(x)}{x}\right| > M\right) \leq \mathbb{P}\left(\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| > M\right) \tag{2.2.33}$$

$$+\,\mathbb{P}\left(\sup_{x\in(\alpha_0,1)} \left|\frac{F_n(x)}{x}\right| > M\right). \tag{2.2.34}$$

Notice that for all $x \in (\alpha_0, 1)$ we surely have

$$\left|\frac{F_n(x)}{x}\right| \leq \frac{1}{\alpha_0}. \tag{2.2.35}$$

Hence if $M > \frac{1}{\alpha_0}$, we have

$$\mathbb{P}\left(\sup_{x\in(\alpha_0,1)} \left|\frac{F_n(x)}{x}\right| > M\right) = 0. \tag{2.2.36}$$

Thus

$$\mathbb{P}\left(\sup_{x\in(0,1)} \left|\frac{F_n(x)}{x}\right| > M\right) \leq \mathbb{P}\left(\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| > M\right) + \mathbf{1}_{\frac{1}{\alpha_0}>M}. \tag{2.2.37}$$

We want to show that for all $\epsilon > 0$ there exist $M > 0$ and $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\mathbb{P}\left(\sup_{x\in(0,1)} \left|\frac{F_n(x)}{x}\right| > M\right) < \epsilon. \tag{2.2.38}$$

If we can instead find $M_1 > 0$ and $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\mathbb{P}\left(\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| > M_1\right) < \epsilon, \tag{2.2.39}$$

we can take $M = \max\{M_1, \frac{1}{\alpha_0} + 1\}$. The result would then immediately follow. Hence we just need to show that

$$\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| = O_p(1), \tag{2.2.40}$$

as $n \to \infty$. Notice that this statement is not dependent on the behaviour of $F_n(x)$ for $x > \alpha_0$. We want to look at the eCDF of a uniform distribution instead of F_n. To do this we define the sequence of random variables

$$Y_i := \begin{cases} \beta_0 X_i, & \text{if } X_i \leq \alpha_0, \\ Z_i, & \text{if } X_i > \alpha_0, \end{cases} \tag{2.2.41}$$

for $i \in \mathbb{N}$, where all Z_i are i.i.d and distributed as a uniform distribution on the interval $[\theta_0, 1]$. Then for any $i \in \mathbb{N}$ and $x \in [0, \theta_0]$ we have

$$\mathbb{P}(Y_i \leq x) = \mathbb{P}\left(X_i \leq \frac{x}{\beta_0}\right) = x. \tag{2.2.42}$$

For any $i \in \mathbb{N}$ and $x \in [\theta_0, 1]$ we have

$$\begin{aligned}
\mathbb{P}(Y_i \leq x) &= \mathbb{P}(Y_i \leq x | X_i \leq \alpha_0)\mathbb{P}(X_i \leq \alpha_0) + \mathbb{P}(Y_i \leq x | X_i > \alpha_0)\mathbb{P}(X_i > \alpha_0) && (2.2.43)\\
&= \theta_0 + \mathbb{P}(Z_i \leq x) \cdot (1 - \theta_0) && (2.2.44)\\
&= \theta_0 + \frac{x - \theta_0}{1 - \theta_0} \cdot (1 - \theta_0) && (2.2.45)\\
&= x. && (2.2.46)
\end{aligned}$$

Therefore $Y_i \sim \text{unif}([0, 1])$. Define G_n to be the eCDF for the sequence of random variables $(Y_1, Y_2, \ldots)$. If $x \leq \alpha_0$, then

$$X_i \leq x \implies Y_n = \beta_0 X_i \leq \beta_0 x, \tag{2.2.47}$$

and

$$Y_i \leq \beta_0 x \implies X_i = \frac{Y_i}{\beta_0} \leq x. \tag{2.2.48}$$

Hence for $x \leq \alpha_0$ we have $F_n(x) = G_n(\beta_0 x)$. Therefore

$$\begin{aligned}
\sup_{x\in(0,\alpha_0)} \left|\frac{F_n(x)}{x}\right| &= \beta_0 \sup_{x\in(0,\alpha_0)} \left|\frac{G_n(\beta_0 x)}{\beta_0 x}\right| && (2.2.49)\\
&= \beta_0 \sup_{x\in(0,\theta_0)} \left|\frac{G_n(x)}{x}\right| && (2.2.50)\\
&\leq \beta_0 \sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right|. && (2.2.51)
\end{aligned}$$

Hence we only need to prove that

$$\sup_{x\in(0,1)} \left| \frac{G_n(x)}{x} \right| = O_p(1). \tag{2.2.52}$$

In Lemma 2.5 it will be shown that this is true. At this moment we shall use this result, but it shall be proven later. From this we can conclude that indeed the first condition is met. Finally we need to show that the second condition is met. First we notice that

$$\sup_{x\in(0,1)} \frac{1-F_n(x)}{1-x} = \sup_{x\in(0,1)} \frac{1-F_n(1-x)}{x}. \tag{2.2.53}$$

We can think of $1 - F_n(1-x)$ as the eCDF of

$$Z_i := 1 - X_i. \tag{2.2.54}$$

But Z_i is distributed with PDF $f(x, 1-\alpha_0, 1-\theta_0)$. Hence Z_i lies within the same family as X_i. If we define G_n to be the eCDF for Z_i, we then have

$$\sup_{x\in(0,1)} \frac{1-F_n(x)}{1-x} = \sup_{x\in(0,1)} \frac{G_n(x)}{x}. \tag{2.2.55}$$

The proof of the fact that the first condition is met by F_n does not depend on α_0 and θ_0 except for the assumption that $\alpha_0 \neq \theta_0$, which still holds for $1-\alpha_0$ and $1-\theta_0$. Hence it can also be applied to G_n, from which we can conclude that the second condition is also satisfied by F_n. Now we have shown that all three conditions of Lemma 2.4 are met by F_n. Hence $\hat{\alpha}_n = \Phi(F_n) = \alpha_0 + o_p(1)$ for $n \to \infty$, which tells us that $\hat{\alpha}_n$ is indeed consistent. ■

2.2.3 Proof of Lemma 2.3

In the proof of Theorem 2.2, we used Lemma 2.3. We will now prove this lemma after giving a heuristic overview of the proof and introducing a relevant lemma.

Recall that Lemma 2.3 states that whenever $G_n(x)$ is a sequence of nonrandom CDFs such that

$$\sup_{x\in(0,1)} \left| \frac{G_n(x)}{x} \right| = O(1), \tag{2.2.56}$$

$$\sup_{x\in(0,1)} \left| \frac{1-G_n(x)}{1-x} \right| = O(1), \tag{2.2.57}$$

and

$$\sup_{x\in(0,1)} |G_n(x) - F_0(x)| = o(1), \tag{2.2.58}$$

where o and O are for $n \to \infty$, then

$$\Phi(G_n) = \alpha_0 + o(1), \tag{2.2.59}$$

for $n \to \infty$. We start by introducing some notation that will be used in this proof. Define $H : [0,1] \times (0,\infty) \times (0,\infty) \to \mathbb{R}$ by

$$H(y_1, y_2, y_3) = y_1 \log y_2 + (1 - y_1) \log y_3. \tag{2.2.60}$$

For all $n \in \mathbb{N}$ define

$$H_n(x) = H\left(G_n(x), \frac{G_n}{x}, \frac{1-G_n}{1-x}\right). \tag{2.2.61}$$

Then by the definition of Φ in Equation (2.2.15), we have

$$\Phi(G_n) = \underset{x\in(0,1)}{\arg\max}\, H_n(x). \tag{2.2.62}$$

Define

$$H_0(x) := F_0(x) \log\left(\frac{F_0(x)}{x}\right) + (1 - F_0(x)) \log\left(\frac{1-F_0(x)}{1-x}\right). \tag{2.2.63}$$

Then

$$\Phi(F_0) = \underset{x\in(0,1)}{\arg\max}\, H_0(x). \tag{2.2.64}$$

We want that $H_n \to H_0$ uniformly. To do this, we need to do a couple of things. When talking about H_n and H_0, we want to be able to consider H on a compact domain for n large enough to exploit uniform continuity with the use of the Heine-Cantor theorem. At the same time, we need be able to consider H_n on a domain that is smaller than $[0,1]$ in order for $G_n(x)/x$ and $(1 - G_n(x))/(1 - x)$ to converge uniformly. This property is necessary for finding a compact domain. To prove that we can use a smaller domain, we need to prove that the argmax of G_n does not converge to 0 or 1. If we have these things, we can show that H_n converges uniformly to H_0 on a closed interval smaller than $[0,1]$ which includes the argmax of H_n for n large enough. Uniform convergence of H_n will allow us to show the final result.

First we bound the domain of the last two parameters of H from above. By Equations (2.2.56) and (2.2.57), we know that

$$\sup_{x\in(0,1)} \frac{G_n(x)}{x} \quad \text{and} \quad \sup_{x\in(0,1)} \frac{G_n(x)}{x}, \tag{2.2.65}$$

are stochastically bounded. It can also be seen that

$$\sup_{x\in(0,1)} \frac{F_0(x)}{x} \quad \text{and} \quad \sup_{x\in(0,1)} \frac{F_0(x)}{x}, \tag{2.2.66}$$

are bounded. This gives us upper bounds M_1 and M_2 respectively. Now we want a positive lower bound for these variables. For this we need uniform convergence of $G_n(x)/x$

and $(1 - G_n(x))/(1 - x)$. However, to do this we need to restrict the domain of x. Hence we need to show that

$$\frac{1}{\Phi(G_n)} = O(1), \quad \text{and} \quad \frac{1}{1 - \Phi(G_n)} = O(1), \tag{2.2.67}$$

for $n \to \infty$. We only prove the former, as the proof of the latter is completely analogous. To prove this, we do a proof by contradiction. We start with a subsequence such that $\Phi(G_{n_k}) \to 0$. Then we will find that $H_{n_k}(\Phi(G_{n_k})) \to 0$, but $H_{n_k}(\alpha_0) \to H_0(\alpha) > 0$. However, that means that at some point $H_{n_k}(\Phi(G_{n_k})) < H_{n_k}(\alpha_0)$, which is not possible because of the definitions of H_n and Φ. Because of this contradiction we conclude that indeed Equation (2.2.67) holds.

Therefore there exist $d_1, d_2 \in (0, 1)$ with $d_1 < d_2$ such that eventually $\Phi(G_n) \in [d_1, d_2]$. We will thus consider $x \in [d_1, d_2]$. Equation (2.2.58) from the hypothesis tell us that G_n converges to F_0 uniformly. Using this uniform convergence we find that we can consider

$$G_n(x) \in [F_0(d_1)/2, (1 + F_0(d_2))/2]. \tag{2.2.68}$$

Now we will find lower bounds for the last two arguments of H. These are based on the fact that it can be shown that

$$\frac{G_n(x)}{x} \to \frac{F_0(x)}{x}, \quad \text{and} \quad \frac{1 - G_n(x)}{1 - x} \to \frac{1 - F_0(x)}{1 - x}, \tag{2.2.69}$$

uniformly on $[d_1, d_2]$. Since these limits can not get arbitrarily close to zero, we can indeed find m_1 and m_2 such that we can now consider H with domain $D := [F_0(d_1)/2, (1 + F_0(d_2))/2] \times [m_1, M_1] \times [m_2, M_2]$, which is compact. Since H is continuous on D and D is compact, we can apply the Heine-Cantor theorem to find that H is uniformly continuous on D. This can be used to show that $H_n \to H_0$ uniformly on $[d_1, d_2]$. We now want to get to the final conclusion. For this we need the following lemma, which will be proven in the next subsection.

Lemma 2.6. *Under the given definitions of Φ, F_0, and α_0 we have $\Phi(F_0) = \alpha_0$.*

We define ζ to be the maximum of H_0 on $[d_1, d_2]$ outside of a small region around α_0. By Lemma 2.6, we know that $\zeta < H_0(\alpha_0)$. But then the maximum of H_n on $[d_1, d_2]$ outside of a small region around α_0 will go towards ζ the maximum of H_n inside of the small region will go towards $H_0(\alpha_0)$. So for n large enough, the maximum of H_n on $[d_1, d_2]$ will lie in the small area around α_0. Since we can take this area arbitrarily small, we find that $\Phi(G_n) = \alpha_0 + o(1)$ for $n \to \infty$. We will now give a detailed proof.

Proof of Lemma 2.3. We start by bounding the domain of the last two parameters of H from above. We need to make sure F_0 and all G_n for n large enough can be properly defined with the new domain of H. By equations 2.2.56 and 2.2.57, there exist $M_1', M_2' \in \mathbb{R}$ and N_1 such that for all $n > N_1$ we have

$$\sup_{x \in (0,1)} \left| \frac{G_n(x)}{x} \right| \leq M_1, \tag{2.2.70}$$

and

$$\sup_{x\in(0,1)} \left| \frac{1-G_n(x)}{1-x} \right| \leq M_2. \tag{2.2.71}$$

Notice that by the definition of F_0 in Equation (2.0.13) we have

$$\frac{F_0(x)}{x} = \begin{cases} \beta_0, & \text{if } x \in [0,\alpha_0], \\ \frac{1-\gamma_0}{x} + \gamma_0, & \text{if } x \in (\alpha_0, 1], \end{cases} \tag{2.2.72}$$

and

$$\frac{1-F_0(x)}{1-x} = \begin{cases} \frac{1-\beta_0 x}{1-x}, & \text{if } x \in [0,\alpha_0], \\ \gamma_0, & \text{if } x \in (\alpha_0, 1]. \end{cases} \tag{2.2.73}$$

Hence $\frac{F_0(x)}{x}$ and $\frac{1-F_0(x)}{1-x}$ are bounded. Therefore M_1 and M_2 can be chosen such that $M_1 \geq M_1'$ and $M_2 \geq M_2'$, but also

$$\frac{F_0(x)}{x} \leq M_1, \quad \text{and} \quad \frac{1-F_0(x)}{1-x} \leq M_2, \tag{2.2.74}$$

for all $x \in (0,1)$. Then for all $n > N_1$ and all $x \in (0,1)$ we have

$$\frac{G_n(x)}{x} \leq M_1 \quad \text{and} \quad \frac{1-G_n(x)}{1-x} \leq M_2. \tag{2.2.75}$$

We can thus consider H as a function on the domain $[0,1] \times (0, M_1] \times (0, M_2]$. We want to be able to consider H on a compact domain for n large enough. To get a positive lower bound for the last two parameters of H, we want to use uniform convergence of $G_n(x)/x$ and $(1-G_n(x))/(1-x)$. Because $G_n(x)/x$ and $(1-G_n(x))/(1-x)$ do not converge uniformly on $[0,1]$, we want to be able to look at a smaller interval for x. To do this, we will show that

$$\frac{1}{\Phi(G_n)} = O(1), \quad \text{and} \quad \frac{1}{1-\Phi(G_n)} = O(1), \tag{2.2.76}$$

for $n \to \infty$, as that would imply that $\Phi(G_n)$ is bounded away from 0 and 1 eventually.

We start by showing that $1/\Phi(G_n) = O(1)$ for $n \to \infty$. We will do this by contradiction. From the hypothesis that $1/\Phi(G_n) \neq O(1)$ we will create some subsequence such that $H_{n_k}(\Phi(G_{n_k}))$ is bounded from above by a sequence that converges to 0. Then we will show that $H_0(\alpha_0) > 0$. Finally we will combine these to come to a contradiction on the basis that $\Phi(G_{n_k})$ should maximise H_{n_k}.

Suppose

$$\frac{1}{\Phi(G_n)} \neq O(1), \tag{2.2.77}$$

for $n \to \infty$. Then there is a subsequence $(G_{n_k})_{k\in\mathbb{N}}$ of $(G_n)_{n\in\mathbb{N}}$ such that

$$\Phi(G_{n_k}) \to 0. \tag{2.2.78}$$

Let $\epsilon > 0$. Then there exists $K_1 \in \mathbb{N}$ such that for all $k > K_1$ we have

$$\Phi(G_{n_k}) < \frac{\epsilon}{M_1}. \tag{2.2.79}$$

By Equation (2.2.70), there exists K_2 such that for all $k > K_2$

$$\frac{G_{n_k}(\Phi(G_{n_k}))}{\Phi(G_{n_k})} \leq M_1. \tag{2.2.80}$$

Hence for $k > \max\{K_1, K_2\} =: K_0$ we have

$$G_{n_k}(\Phi(G_{n_k})) \leq M_1 \Phi(G_{n_k}) < \epsilon. \tag{2.2.81}$$

Therefore

$$G_{n_k}(\Phi(G_{n_k})) \to 0. \tag{2.2.82}$$

By equations 2.2.78, 2.2.80 and 2.2.82 we find that for all $k > K_2$ we have

$$\begin{aligned} H_{n_k}(\Phi(G_{n_k})) &= G_{n_k}(\Phi(G_{n_k})) \log\left(\frac{G_{n_k}(\Phi(G_{n_k}))}{\Phi(G_{n_k})}\right) & (2.2.83)\\ &+ (1 - G_{n_k}(\Phi(G_{n_k}))) \log\left(\frac{1 - G_{n_k}(\Phi(G_{n_k}))}{1 - \Phi(G_{n_k})}\right) & (2.2.84)\\ &\leq G_{n_k}(\Phi(G_{n_k})) \log M_1 + (1 - G_{n_k}(\Phi(G_{n_k}))) \log\left(\frac{1 - G_{n_k}(\Phi(G_{n_k}))}{1 - \Phi(G_{n_k})}\right) & (2.2.85)\\ &\to 0. & (2.2.86) \end{aligned}$$

Therefore

$$\forall_{\epsilon>0} \exists_{K \in \mathbb{N}} : \forall_{k \geq K} : H_{n_k}(\Phi(G_{n_k})) < \epsilon. \tag{2.2.87}$$

We will now show that $H_0(\alpha_0) > 0$. By the definitions of H_0 and F_0 given in Equation (2.2.63) and Equation (2.0.13) respectively, we know that

$$H_0(\alpha_0) = \theta_0 \log \frac{\theta_0}{\alpha_0} + (1 - \theta_0) \log \frac{1 - \theta_0}{1 - \alpha_0}. \tag{2.2.88}$$

Define $J : (0, 1) \times (0, 1) \to \mathbb{R}$ by

$$J(\alpha, \theta) = \theta \log \frac{\theta}{\alpha} + (1 - \theta) \log \frac{1 - \theta}{1 - \alpha}. \tag{2.2.89}$$

Its gradient is given by

$$\nabla J(\alpha, \theta) = \begin{pmatrix} \frac{1-\theta}{1-\alpha} - \frac{\theta}{\alpha} \\ \log \frac{\theta}{1-\theta} + \log \frac{1-\alpha}{\alpha} \end{pmatrix}, \tag{2.2.90}$$

which can easily be seen to be zero if and only if $\alpha = \theta$. When $\alpha = \theta$, we have

$$J(\alpha, \alpha) = 0. \tag{2.2.91}$$

Since

$$J\left(\frac{1}{4},\frac{1}{2}\right) = J\left(\frac{3}{4},\frac{1}{2}\right) = \frac{1}{2}\log 2 + \frac{1}{2}\log\frac{4}{6} = \frac{1}{2}\log\frac{4}{3} > 0, \tag{2.2.92}$$

we find that $J(\alpha,\theta) > 0$ on both sides of the line $\alpha = \theta$. Therefore $J(\alpha,\theta) > 0$ whenever $\alpha \neq \theta$. Since $\alpha_0 \neq \theta_0$ by Assumption 2.1, we have $H_0(\alpha_0) = J(\alpha_0,\theta_0) > 0$.

By continuity of H and convergence of G_n to F_0, we know that $H_n \to H_0$ pointwise. Hence $H_{n_k}(\alpha_0) \to H_0(\alpha_0) > 0$. Let $\epsilon = H_0(\alpha_0)/2$. Then there exists $K \in \mathbb{N}$ such that for all $k > K$

$$H_{n_k}(\alpha_0) > \frac{1}{2}H_0(\alpha_0). \tag{2.2.93}$$

However, by Equation (2.2.87) there exists $K' \in \mathbb{N}$ such that for $k > K'$ we have

$$H_{n_k}(\Phi(G_{n_k})) < \frac{1}{2}H_0(\alpha_0). \tag{2.2.94}$$

But then for $k > \max\{K, K'\}$ we have

$$H_{n_k}(\Phi(G_n)) < H_{n_k}(\alpha_0), \tag{2.2.95}$$

which contradicts that $\Phi(G_{n_k})$ maximises H_{n_k}. We can thus conclude that

$$\frac{1}{\Phi(G_n)} = O(1). \tag{2.2.96}$$

In an analogous way, it can be shown that

$$\frac{1}{1-\Phi(G_n)} = O(1). \tag{2.2.97}$$

Hence there exist $d_1, d_2 \in (0,1)$ with $d_1 < d_2$ and $N_2 \in \mathbb{N}$ such that for all $n > N_2$ we have

$$\Phi(G_n) \in [d_1, d_2]. \tag{2.2.98}$$

We will consider $x \in [d_1, d_2]$. Recall that F_0 is increasing, because it is a CDF. Equation (2.2.58) from the lemma hypothesis tells us that G_n converges uniformly to F_0. By using uniform convergence of G_n to F_0, we know that we can choose $N_3 \in \mathbb{N}$ such that for all $n > N_3$ and $x \in [d_1, d_2]$ we have

$$G_n(x) \in [F_0(d_1)/2, (1 + F_0(d_2))/2]. \tag{2.2.99}$$

Hence we will consider H on the domain $[F_0(d_1)/2, (1 + F_0(d_2))/2] \times (0, M_1] \times (0, M_2]$, but first we shall make it compact by getting lower bounds for the last two arguments based on the restriction on the first one. This will give us a domain of the form $[F_0(d_1)/2, (1 + F_0(d_2))/2] \times [m_1, M_1] \times [m_2, M_2]$. To do this, we first need to show that

$$\frac{G_n(x)}{x} \to \frac{F_0(x)}{x}, \tag{2.2.100}$$

and

$$\frac{1-G_n(x)}{1-x} \to \frac{1-F_0(x)}{1-x}, \tag{2.2.101}$$

uniformly on $[d_1, d_2]$. We start by proving the former. Let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\sup_{x\in(0,1)} |G_n(x) - F_0(x)| < d_1 \cdot \epsilon. \tag{2.2.102}$$

Then for all $n > N$ and $x \in (0, 1)$ we have

$$|G_n(x) - F_0(x)| < d_1 \cdot \epsilon. \tag{2.2.103}$$

Hence for all $n > N$ and $x \in (0, 1)$ we have

$$\left|\frac{G_n(x)}{x} - \frac{F_0(x)}{x}\right| = \frac{1}{x}|G_n(x) - F_0(x)| \tag{2.2.104}$$

$$\leq \frac{1}{d_1}|G_n(x) - F_0(x)| \tag{2.2.105}$$

$$< \epsilon. \tag{2.2.106}$$

Therefore $G_n(x)/x$ converges to $F_0(x)/x$ uniformly on $[d_1, d_2]$. We now turn towards $(1 - G_n(x))/(1 - x)$. Let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\sup_{x\in(0,1)} |G_n(x) - F_0(x)| < (1 - d_2) \cdot \epsilon. \tag{2.2.107}$$

Then for all $n > N$ and $x \in (0, 1)$ we have

$$|G_n(x) - F_0(x)| < (1 - d_2) \cdot \epsilon. \tag{2.2.108}$$

Hence for all $n > N$ and $x \in (0, 1)$ we have

$$\left|\frac{1-G_n(x)}{1-x} - \frac{1-F_0(x)}{1-x}\right| = \frac{1}{1-x}|G_n(x) - F_0(x)| \tag{2.2.109}$$

$$\leq \frac{1}{1-d_2}|G_n(x) - F_0(x)| \tag{2.2.110}$$

$$< \epsilon. \tag{2.2.111}$$

Therefore $(1 - G_n(x))/(1 - x)$ converges to $(1 - F_0(x))/(1 - x)$ uniformly on $[d_1, d_2]$. We will now find m_1, m_2 and N_3 such that for all $n > N_3$ and $x \in [d_1, d_2]$ we have

$$\frac{G_n(x)}{x} > m_1, \tag{2.2.112}$$

and

$$\frac{1-G_n(x)}{1-x} > m_2. \tag{2.2.113}$$

By the definition of F_0 given in Equation (2.0.13), we know that

$$\frac{F_0(x)}{x} = \begin{cases} \beta_0, & \text{if } x \in [0, \alpha_0], \\ \frac{1-\gamma_0}{x} + \gamma_0, & \text{if } x \in (\alpha_0, 1]. \end{cases} \tag{2.2.114}$$

Note that $\beta_0 > 0$. Since $\gamma_0 \in (0, 1)$, we have

$$\frac{1-\gamma_0}{x} + \gamma_0 > 0, \tag{2.2.115}$$

for all $x \in (\alpha_0, 1]$. Hence in neither of the cases for x in Equation (2.2.114), can $F_0(x)/x$ get arbitrarily close to 0. Similarly we know that

$$\frac{1-F_0(x)}{1-x} = \begin{cases} \frac{1-\beta_0 x}{1-x}, & \text{if } x \in [0, \alpha_0], \\ \gamma_0, & \text{if } x \in (\alpha_0, 1]. \end{cases} \tag{2.2.116}$$

Note that $\gamma_0 > 0$. Since $\beta_0 = \theta_0/\alpha_0$, we have

$$1 - \beta_0 x = 1 - \frac{\theta_0 x}{\alpha_0} > 0, \tag{2.2.117}$$

for $x \in [0, \alpha_0)$. Therefore

$$\frac{1-\beta_0 x}{1-x} \geq 1 - \beta_0 x > 0, \tag{2.2.118}$$

for $x \in [0, \alpha_0)$. Hence $\frac{1-F_0(x)}{1-x}$ can not get arbitrarily close to 0. Let

$$m_3 := \min_{x \in [d_1, d_2]} \frac{F_0(x)}{x} > 0, \tag{2.2.119}$$

and

$$m_4 := \min_{x \in [d_1, d_2]} \frac{1-F_0(x)}{1-x} > 0. \tag{2.2.120}$$

By the uniform convergence on $[d_1, d_2]$ mentioned in equations 2.2.100 and 2.2.101, there exists $N_4 \in \mathbb{N}$ such that for all $n > N_4$ and $x \in [d_1, d_2]$ we have

$$\frac{G_n(x)}{x} > m_1 := \frac{m_3}{2}, \tag{2.2.121}$$

and

$$\frac{1-G_n(x)}{1-x} > m_2 := \frac{m_4}{2}. \tag{2.2.122}$$

Notice that we also have

$$\frac{F_0(x)}{x} \geq m_1 \quad \text{and} \quad \frac{1-F_0(x)}{1-x} \geq m_2, \tag{2.2.123}$$

for all $x \in [d_1, d_2]$. We will be able to consider H on the domain $D := [F_0(d_1)/2, (1 + F_0(d_2))/2] \times [m_1, M_1] \times [m_2, M_2]$. This domain is compact. Since H is continuous on D and D is compact, we can apply the Heine-Cantor theorem to find that H is uniformly continuous on D. Let $\epsilon > 0$. By uniform continuity of H on D, there exists $\delta > 0$ such that for all $y_1, y_2 \in D$ we have

$$|y_1 - y_2| < \delta \implies |H(y_1) - H(y_2)| < \epsilon. \tag{2.2.124}$$

Let $N_5 \in \mathbb{N}$ such that for all $n > N_5$ and $x \in (0, 1)$ we have

$$|G_n(x) - F_0(x)| < \frac{1}{1 + \frac{1}{d_1} + \frac{1}{1-d_2}} \cdot \delta. \tag{2.2.125}$$

Let $N_0 := \max\{N_1, \ldots, N_5\}$ and let $x \in [d_1, d_2]$ be arbitrary. Define

$$x_n := \begin{pmatrix} G_n(x) \\ \frac{G_n(x)}{x} \\ \frac{1-G_n(x)}{1-x} \end{pmatrix}, \tag{2.2.126}$$

and

$$y := \begin{pmatrix} F_0(x) \\ \frac{F_0(x)}{x} \\ \frac{1-F_0(x)}{1-x} \end{pmatrix}. \tag{2.2.127}$$

By definition of $N_1, \ldots, N_5$, we know that for $n > N_0$ we have $x_n, y \in D$ and

$$\begin{aligned} |x_n - y| &\leq |G_n(x) - F_0(x)| + \left|\frac{G_n(x)}{x} - \frac{F_0(x)}{x}\right| + \left|\frac{1 - G_n(x)}{1-x} - \frac{1 - F_0(x)}{1-x}\right| && (2.2.128) \\ &= \left(1 + \frac{1}{x} + \frac{1}{1-x}\right) |G_n(x) - F_0(x)| && (2.2.129) \\ &\leq \left(1 + \frac{1}{d_1} + \frac{1}{1-d_2}\right) |G_n(x) - F_0(x)| && (2.2.130) \\ &< \delta. && (2.2.131) \end{aligned}$$

By Equation (2.2.124) we thus have

$$|H(x_n) - H(y)| < \epsilon. \tag{2.2.132}$$

Therefore, for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $x \in D$ and $n > N$ we have

$$|H_n(x) - H_0(x)| = |H(x_n) - H(y)| < \epsilon. \tag{2.2.133}$$

Hence $H_n(x) \to H_0(x)$ uniformly on $[d_1, d_2]$. We finally turn ourselves to $\Phi(G_n)$. Let $\epsilon > 0$ and define

$$\zeta := \max_{x \in [d_1,d_2] \setminus (\alpha_0 - \epsilon, \alpha_0 + \epsilon)} H_0(x), \tag{2.2.134}$$

and

$$\delta = \frac{1}{2}(H_0(\alpha_0) - \zeta). \tag{2.2.135}$$

By Lemma 2.6 we have

$$\zeta < H_0(\alpha_0), \tag{2.2.136}$$

and thus $\delta > 0$. By uniform convergence of H_n to H_0 on $[d_1, d_2]$, we can choose $N \in \mathbb{N}$ such that for all $x \in [d_1, d_2]$ and $n > N$ we have

$$|H_n(x) - H_0(x)| < \delta. \tag{2.2.137}$$

Let $n > N$. Then we find that

$$\max_{x \in [d_1,d_2] \setminus (\alpha_0 - \epsilon, \alpha_0 + \epsilon)} H_n(x) < \zeta + \delta = \frac{1}{2}(H_0(\alpha_0) + \zeta), \tag{2.2.138}$$

and

$$\max_{x \in (\alpha_0 - \epsilon, \alpha_0 + \epsilon)} H_n(x) > H_0(\alpha_0) - \delta = \frac{1}{2}(H_0(\alpha_0) + \zeta). \tag{2.2.139}$$

Since $\Phi(G_n) \in [d_1, d_2]$, we thus have

$$|\Phi(G_n) - \alpha_0| < \epsilon. \tag{2.2.140}$$

We conclude that

$$\Phi(G_n) = \alpha_0 + o(1), \tag{2.2.141}$$

for $n \to \infty$. ■

2.2.4 Proof of Lemma 2.6

To complete this section, we now only need to prove Lemma 2.6, which states that $\Phi(F_0) = \alpha_0$.

Proof of Lemma 2.6. We use the same notation as in the previous subsection. In order to prove that $\alpha_0 = \Phi(F_0)$, we need to show that α_0 maximises H_0 on $(0, 1)$. Usually we would like to simply find where the derivative of H_0 is zero. However, F_0 is not differentiable in α_0, leading us to a different approach. We will show that $H_0'(x) < 0$ for all $x \in (\alpha_0, 1)$ and $H_0'(x) > 0$ for all $x \in (0, \alpha_0)$. If we also have that H_0 is continuous, we will be able to conclude that indeed $\alpha_0 = \Phi(F_0)$.

Notice that F_0 is continuous. Therefore H_0 is also continuous. When proving properties of the derivative of H_0, it will be useful to extend the domain of H_0 to $[0, 1]$. We can extend the domain of H_0 by setting $H_0(0) = \lim_{x\to0} H_0(x)$ and $H_0(1) = \lim_{x\to1} H_0(x)$. This will then maintain continuity. Now H_0 is a continuous function with domain $[0, 1]$.

First we note that

$$H_0(\alpha_0) = \beta_0 \alpha_0 \log \beta_0 + \gamma_0 (1 - \alpha_0) \log \gamma_0, \tag{2.2.142}$$

is well defined. We will first look at $(0, \alpha_0)$. For $x \in (0, \alpha_0)$ we have

$$H_0(x) = \beta_0 x \log \frac{\beta_0 x}{x} + (1 - \beta_0 x) \log \frac{1 - \beta_0 x}{1 - x} \tag{2.2.143}$$

$$= \beta_0 x \log \beta_0 + (1 - \beta_0 x) \log \frac{1 - \beta_0 x}{1 - x}. \tag{2.2.144}$$

Since this is continuous, it also holds for $x = 0$. Therefore $H_0(0) = 0 < H_0(\alpha_0)$. By straightforward differentiation we find that for $x \in [0, \alpha_0)$ we have

$$H_0'(x) = \beta_0 \log \beta_0 + (1 - \beta_0 x) \left(\frac{1}{1 - x} - \frac{\beta_0}{1 - \beta_0 x} \right) - \beta_0 \log \frac{1 - \beta_0 x}{1 - x}. \tag{2.2.145}$$

Since $\theta_0 \in (0, 1)$ and $x \in [0, \alpha_0)$, we have

$$1 - \beta_0 x = 1 - \frac{\theta_0}{\alpha_0} x > 0. \tag{2.2.146}$$

Hence we can write

$$H_0'(x) = \beta_0 \log \beta_0 + \frac{1 - \beta_0 x}{1 - x} - \beta_0 - \beta_0 \log \frac{1 - \beta_0 x}{1 - x}, \tag{2.2.147}$$

for $x \in [0, \alpha_0)$ and this derivative is well defined. We will show that this is non-negative for all $x \in [0, \alpha_0)$. To do this we will show that the value for $x = 0$ is non-negative and that the second derivative is non-negative for $x \in (0, \alpha_0)$. First we look at $H_0'(0)$. We see that this is

$$H_0'(0) = 1 - \beta_0 + \beta_0 \log \beta_0. \tag{2.2.148}$$

To show that this is positive, we consider $\omega : (0, \infty) \to \mathbb{R}$ defined by

$$\omega(\beta) := 1 - \beta + \beta \log \beta. \tag{2.2.149}$$

Then

$$\frac{d\omega}{d\beta} = \log \beta, \tag{2.2.150}$$

which is negative for $\beta \in (0, 1)$ and positive for $\beta > 1$. Its only zero lies at $\beta = 1$. Since $\omega(1) = 0$ and ω is continuous, we find that $\omega(\beta) > 0$ for all $\beta \in (0, \infty)$. By Assumption 2.1, we have $\beta_0 \neq 1$. Hence, no matter what the value for β_0 is, we always have $H_0'(0) > 0$. We turn ourselves towards the second derivative of H_0 for $x \in (0, \alpha_0)$. By differentiating Equation (2.2.147) we find that the second derivative of H_0 for $x \in (0, \alpha_0)$ is given by

$$H_0''(x) = \frac{-\beta_0(1 - x) + (1 - \beta_0 x)}{(1 - x)^2} + \frac{\beta_0^2}{1 - \beta_0 x} - \frac{\beta_0}{1 - x} \tag{2.2.151}$$

$$= \frac{(\beta_0 - 1)^2}{(1 - x)^2 (1 - \beta_0 x)}. \tag{2.2.152}$$

By Equation (2.2.146) we have $(1-\beta_0 x) > 0$. Since $x \neq 1$ we have $(1-x)^2 > 0$. By Assumption 2.1 we have $\beta_0 \neq 1$ and therefore $(\beta_0 - 1)^2 > 0$. Thus $H''(x) > 0$ for all $x \in (0, \alpha_0)$ and $H_0'(0) > 0$. By continuity of $H_0'(x)$ on $[0, \alpha_0)$ we find that $H'(x) > 0$ for all $x \in [0, \alpha_0)$.

We turn ourselves to $H_0(x)$ for $x \in (\alpha_0, 1)$. For $x \in (\alpha_0, 1)$ we have

$$H_0(x) = (1-\gamma_0+\gamma_0 x)\log\frac{1-\gamma_0+\gamma_0 x}{x} + \gamma_0 \log\gamma_0 - x\gamma_0\log\gamma_0. \tag{2.2.153}$$

By continuity, this also holds for $x = 1$. Taking the derivative yields

$$H_0'(x) = (1-\gamma_0+\gamma_0 x)\left(\frac{\gamma_0}{1-\gamma_0+\gamma_0 x} - \frac{1}{x}\right) + \gamma_0\log\frac{1-\gamma_0+\gamma_0 x}{x} - \gamma_0\log\gamma_0, \tag{2.2.154}$$

for $x > \alpha_0$. Since $\theta_0 \in (0,1)$ and $x > \alpha_0$, we have

$$1 - \gamma_0(1-x) > 0, \tag{2.2.155}$$

Hence for $x \in (\alpha_0, 1]$ we can write

$$H_0'(x) = \gamma_0 - \frac{1-\gamma_0+\gamma_0 x}{x} + \gamma_0\log\frac{1-\gamma_0+\gamma_0 x}{x} - \gamma_0\log\gamma_0, \tag{2.2.156}$$

and this derivative is well-defined. We want to show that $H_0'(x) < 0$ for all $x \in (\alpha_0, 1]$. To do this we will show that $H_0'(1) < 0$ and $H_0''(x) > 0$ for $x \in (\alpha_0, 1)$. The result then follows from continuity of $H_0'(x)$ on $(\alpha_0, 1]$. First we notice that

$$H_0'(1) = \gamma_0 - 1 - \gamma_0\log\gamma_0 = -\omega(\gamma_0), \tag{2.2.157}$$

where ω is as defined in Equation (2.2.149). We showed that ω is a positive function on $(0,\infty)\setminus\{1\}$. By Assumption 2.1 we know that $\gamma_0 \neq 1$. It follows that $H_0'(1) < 0$. We continue by looking at $H_0''(x)$. For $x \in (\alpha_0, 1)$ we have

$$H_0''(x) = \frac{1-\gamma_0}{x^2} + \frac{\gamma_0^2}{1-\gamma_0(1-x)} - \frac{\gamma_0}{x} \tag{2.2.158}$$

$$= \frac{(\gamma_0-1)^2}{x^2(1-\gamma_0(1-x))}. \tag{2.2.159}$$

By Equation (2.2.155) we have $1-\gamma_0(1-x) > 0$. Since $x > \alpha_0$, we have $x^2 > 0$. Since by Assumption 2.1 we know that $\gamma_0 \neq 1$, we have $(\gamma_0 - 1)^2 > 0$. Thus $H''(x) > 0$ for all $x \in (\alpha_0, 0)$ and $H_0'(1) < 0$. By continuity of $H_0'(x)$ on $(\alpha_0, 1]$ we find that $H'(x) < 0$ for all $x \in (\alpha_0, 1]$.

Now we have $H_0'(x) < 0$ for all $x \in (\alpha_0, 1]$ and $H_0'(x) > 0$ for all $x \in [0, \alpha_0)$. By continuity of $H_0(x)$ we can conclude that indeed $\Phi(F_0) = \alpha_0$. ■

2.3 Quasi-maximum likelihood estimator

Our goal in this chapter is to show that $\hat{\alpha}_n = \alpha_0 + O_p(1/n)$. Instead of showing this directly, we will introduce an estimator for α_0 in this section that will be close enough to $\hat{\alpha}_n$, but also close enough to α_0. To do this we will use a quasi-maximum likelihood estimator (QMLE). A QMLE is an estimator which maximises a function that is related to the log-likelihood. To create such a QMLE, we will rewrite the log-likelihood divided by n to include terms that have convenient asymptotic properties. We will then use one part of this resulting function. The QMLE will maximise this function in some small neighbourhood around α_0. We start out by displaying the QMLE and then proceed by motivating the choice for this estimator. First notice that we can use the actual values for α_0 and θ_0 in the definition of the QMLE, because we only use the QMLE as a vehicle for proving properties of the MLE. We now need to define two things that are necessary for the definition of the QMLE. First will be a function that will be maximised in some range around α_0 by the QMLE and the second will be a sequence of ranges around α_0 that the QMLE will maximise this function on.

The function that we want to use for defining the QMLE is denoted by

$$M_n(\alpha, \theta) := F_n(\alpha) - F_n(\alpha_0) + \frac{(\gamma_0 - \beta_0)(\alpha - \alpha_0)}{\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha_0, \theta))}. \tag{2.3.1}$$

To show that this function makes sense, we will state a theorem that tells us that the log-likelihood can be expressed in some way including asymptotically small terms. These terms are small O or small O in probability. The deterministic small Os are for $\alpha \to \infty$. The probabilistic equivalents are for $n \to \infty$. However, sometimes a term is a function of both n and α. We want to know what happens when $n \to \infty$ and $\alpha \to \alpha_0$. For this we need to introduce new notation, as the big O small O notation does not include enough rigour when it comes to this multivariate asymptotic behaviour. When we will be using these expressions in proof later on, we will be evaluating in consistent estimators for α_0. Therefore we only care about what happens when we evaluate the term in a consistent estimator for α_0 and then take n to infinity. We use the following notation.

Definition 2.1. Let $g_n : \mathbb{R} \to \mathbb{R}$ be a sequence of random functions. Let $h : \mathbb{R} \to \mathbb{R}$. Fix some $a \in \mathbb{R}$. Then we write

$$g_n(\alpha) = T_p(h(\alpha)), \tag{2.3.2}$$

for $\alpha \to a$ if for any consistent estimator η_n of a we have

$$g_n(\eta_n) = o_p(h(\eta_n)), \tag{2.3.3}$$

for $n \to \infty$.

Now we can state the theorem using this new notation.

Theorem 2.7. *The log-likelihood divided by n has the property that*

$$\ell_n(\alpha, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log(f(X_i, \alpha_0, \theta)) + S_n(\alpha, \theta), \tag{2.3.4}$$

for some function

$$\begin{aligned} S_n(\alpha, \theta) = (\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha_0, \theta)) + o(1)) M_n(\alpha, \theta) \\ + (\alpha - \alpha_0)(o(1) + o_p(1)) + T_p(|\alpha - \alpha_0|), \end{aligned} \tag{2.3.5}$$

where o and T_p are for $\alpha \to \alpha_0$ and o_p is for $n \to \infty$.

Most terms in this expression are asymptotically small. Because of Assumption 2.2, which states that $\beta_0 > \gamma_0$, we also know that $\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha_0, \theta)) + o(1)$ will be positive when θ is close enough to θ_0 and α is close enough to α_0. Hence it does seem to make sense to maximise $M_n(\alpha, \theta_0)$ in some region around α_0.

To create a sequence of ranges around α_0 that the QMLE will maximise M_n on, we will use the following lemma that shall be proven after defining the QMLE.

Lemma 2.8. *Suppose $\hat{\eta}_n$ is a consistent estimator of some parameter η_0. Then there exists a sequence $(a_n)_{n \in \mathbb{N}}$ of positive real numbers with $\lim_{n \to \infty} a_n = 0$ such that*

$$\hat{\eta}_n - \eta_0 = o_p(a_n), \tag{2.3.6}$$

as $n \to \infty$.

It has been proven that $\hat{\alpha}_n$ is a consistent estimator of α_0. By Lemma 2.8 it thus follows that we can define a sequence a_n of positive numbers with $\lim_{n \to \infty} a_n = 0$ such that

$$\hat{\alpha}_n - \alpha_0 = o_p(a_n), \tag{2.3.7}$$

as $n \to \infty$.

We can now define the new estimator $\tilde{\alpha}$ using this sequence a_n and the function $M_n(\alpha, \theta)$.

Definition 2.2. The QMLE $\tilde{\alpha}_n$ for α_0 is defined as

$$\tilde{\alpha}_n := \underset{\alpha \in [\alpha_0 - a_n, \alpha_0 + a_n]}{\arg\max} M_n(\alpha, \theta_0). \tag{2.3.8}$$

Now we move on to the proof of Lemma 2.8.

Proof of Lemma 2.8. By consistency of $\hat{\eta}_n$ we know that for each $\epsilon, \delta > 0$ there exists $N_{\epsilon,\delta} \in \mathbb{N}$ such that for all $n > N_{\epsilon,\delta}$ we have

$$\mathbb{P}(|\hat{\eta}_n - \eta_0| > \delta) < \epsilon. \tag{2.3.9}$$

Let $\epsilon, \delta > 0$ and for each $i \in \mathbb{N}$ define $\delta_i := \delta/i$. Then for each $i \in \mathbb{N}$ there exists some $N_i \in \mathbb{N}$ such that for all $n > N_i$ we have

$$\mathbb{P}(i \cdot |\hat{\eta}_n - \eta_0| > \delta) = \mathbb{P}(|\hat{\eta}_n - \eta_0| > \delta_i) < \epsilon. \tag{2.3.10}$$

Notice that the sequence $(N_i)_{i \in \mathbb{N}}$ can be chosen to be strictly increasing. Then for each $i \in \mathbb{N}$ we can define

$$a_n = \frac{1}{i}, \tag{2.3.11}$$

for $n \in \{N_i + 1 \ldots, N_{i+1}\}$. For $n \leq N_1$, we define $a_n = 1$. Clearly all elements are positive and

$$a_n \xrightarrow{n \to \infty} 0. \tag{2.3.12}$$

Let $n > N_1$. Then

$$\mathbb{P}\left(\frac{|\hat{\eta}_n - \eta_0|}{a_n} > \delta\right) = \mathbb{P}(i \cdot |\hat{\eta}_n - \eta_0| > \delta), \tag{2.3.13}$$

for some $i \in \mathbb{N}$. But then by definition of (a_n) we know that $n > N_i$. Therefore by definition of N_i we have

$$\mathbb{P}(i \cdot |\hat{\eta}_n - \eta_0| > \delta) < \epsilon. \tag{2.3.14}$$

We conclude that

$$\hat{\eta}_n - \eta_0 = o_p\left(a_n\right), \tag{2.3.15}$$

as $n \to \infty$. ■

To finish this section, we will prove Theorem 2.7.

Proof of Theorem 2.7. The likelihood is given by

$$\mathcal{L}_n(\alpha, \theta) = \prod_{i=1}^{n} f(X_i, \alpha, \theta). \tag{2.3.16}$$

Instead we look at the log-likelihood divided by n, which is given by

$$\ell_n(\alpha, \theta) = \frac{1}{n}\sum_{i=1}^{n} \log(f(X_i, \alpha, \theta)) \tag{2.3.17}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \log(f(X_i, \alpha_0, \theta)) + \frac{1}{n}\sum_{i=1}^{n} \big(\log(f(X_i, \alpha, \theta) - \log(f(X_i, \alpha_0, \theta)\big). \tag{2.3.18}$$

The first sum in Equation (2.3.18) does not depend on α. Hence it makes sense to ignore that part when constructing the QMLE for α_0. We will denote the second part of Equation (2.3.18) as

$$S_n(\alpha, \theta) := \frac{1}{n}\sum_{i=1}^{n} \big(\log(f(X_i, \alpha, \theta) - \log(f(X_i, \alpha_0, \theta)\big). \tag{2.3.19}$$

We want to split this sum into two. The first sum will only include the observations that lie between α and α_0. The second sum will include all other observations. To be able to talk about the observations between α and α_0, we denote the minimum and maximum of $\{\alpha, \alpha_0\}$ by

$$\alpha_m(\alpha) := \min\{\alpha, \alpha_0\}, \tag{2.3.20}$$

and

$$\alpha_M(\alpha) := \max\{\alpha, \alpha_0\}, \tag{2.3.21}$$

respectively. Now we can define the two separate sums using this notation. The sum over the observations between α and α_0 is denoted by

$$S_n^{(1)}(\alpha, \theta) := \frac{1}{n} \sum_{X_i \in \left(\alpha_m(\alpha), \alpha_M(\alpha)\right)} \Big(\log(f(X_i, \alpha, \theta)) - \log(f(X_i, \alpha_0, \theta)) \Big), \tag{2.3.22}$$

and the sum over the other observations is denoted by

$$S_n^{(2)}(\alpha, \theta) := \frac{1}{n} \sum_{X_i \notin \left(\alpha_m(\alpha), \alpha_M(\alpha)\right)} \Big(\log(f(X_i, \alpha, \theta)) - \log(f(X_i, \alpha_0, \theta)) \Big). \tag{2.3.23}$$

Now we have $S_n(\alpha, \theta) = S_n^{(1)}(\alpha, \theta) + S_n^{(2)}(\alpha, \theta)$. We will rewrite $S_n^{(1)}$ and $S_n^{(2)}$ separately. We start by looking at $S_n^{(1)}$. First notice that for any observation $X_i \in (\alpha_m(\alpha), \alpha_M(\alpha))$

$$\log(f(X_i, \alpha, \theta)) - \log(f(X_i, \alpha_0, \theta)) = \begin{cases} \log(\beta(\alpha, \theta)) - \log(\gamma(\alpha_0, \theta)), & \text{if } \alpha > \alpha_0, \\ \log(\gamma(\alpha, \theta)) - \log(\beta(\alpha_0, \theta)), & \text{if } \alpha < \alpha_0. \end{cases} \tag{2.3.24}$$

Recall the definition of the eCDF from Equation (2.0.14). Then using Equation (2.3.24) we find that

$$S_n^{(1)}(\alpha, \theta) = \begin{cases} (F_n(\alpha) - F_n(\alpha_0))\Big(\log(\beta(\alpha, \theta)) - \log(\gamma(\alpha_0, \theta)) \Big), & \text{if } \alpha > \alpha_0, \\ (F_n(\alpha) - F_n(\alpha_0))\Big(\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha, \theta)) \Big), & \text{if } \alpha < \alpha_0. \end{cases} \tag{2.3.25}$$

We can rewrite this as

$$S_n^{(1)}(\alpha, \theta) = (F_n(\alpha) - F_n(\alpha_0))\Big(\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha_0, \theta)) + \xi(\alpha, \theta) \Big), \tag{2.3.26}$$

where

$$\xi(\alpha, \theta) := \begin{cases} \log(\beta(\alpha, \theta)) - \log(\beta(\alpha_0, \theta)), & \text{if } \alpha > \alpha_0, \\ \log(\gamma(\alpha_0, \theta)) - \log(\gamma(\alpha, \theta)), & \text{if } \alpha < \alpha_0. \end{cases} \tag{2.3.27}$$

Then

$$\begin{aligned} S_n^{(1)}(\alpha,\theta) &= (F_n(\alpha) - F_n(\alpha_0))\log\left(\frac{\beta(\alpha_0,\theta)}{\gamma(\alpha_0,\theta)}\right) \\ &+ \xi(\alpha,\theta)\left(F_n(\alpha) - F_n(\alpha_0) + \frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta)) - \log(\gamma(\alpha_0,\theta))}\right) \qquad (2.3.28) \\ &- \xi(\alpha,\theta)\frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta)) - \log(\gamma(\alpha_0,\theta))} \\ &= (F_n(\alpha) - F_n(\alpha_0))\log\left(\frac{\beta(\alpha_0,\theta)}{\gamma(\alpha_0,\theta)}\right) + \xi(\alpha,\theta)M_n(\alpha,\theta) \\ &- \xi(\alpha,\theta)\frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta)) - \log(\gamma(\alpha_0,\theta))}. \qquad (2.3.29) \end{aligned}$$

Notice that both parts of $\xi(\alpha,\theta)$ are continuous in α. Hence for all $\theta \in (0,1)$ we have

$$\lim_{\alpha\downarrow\alpha_0} \xi(\alpha,\theta) = \lim_{\alpha\downarrow\alpha_0} \log(\beta(\alpha,\theta)) - \log(\beta(\alpha_0,\theta)) = 0, \qquad (2.3.30)$$

and

$$\lim_{\alpha\uparrow\alpha_0} \xi(\alpha,\theta) = \lim_{\alpha\uparrow\alpha_0} \log(\gamma(\alpha_0,\theta)) - \log(\gamma(\alpha,\theta)) = 0. \qquad (2.3.31)$$

Therefore

$$\lim_{\alpha\to\alpha_0} \xi(\alpha,\theta) = 0, \qquad (2.3.32)$$

for any $\theta \in (0,1)$. Thus

$$\xi(\alpha,\theta) = o(1), \qquad (2.3.33)$$

for $\alpha \to \alpha_0$. Therefore

$$-\xi(\alpha,\theta)\frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta)) - \log(\gamma(\alpha_0,\theta))} = (\alpha-\alpha_0)o(1), \qquad (2.3.34)$$

for $\alpha \to \alpha_0$. Plugging this back into Equation (2.3.29) yields

$$S_n^{(1)}(\alpha,\theta) = (F_n(\alpha) - F_n(\alpha_0))\log\left(\frac{\beta(\alpha_0,\theta)}{\gamma(\alpha_0,\theta)}\right) + o(1)M_n(\alpha,\theta) + o(1)(\alpha-\alpha_0), \qquad (2.3.35)$$

for $\alpha \to \alpha_0$.

We now turn ourselves towards $S_n^{(2)}(\alpha,\theta)$. Since we have

$$\mathbb{P}(X_i \in \{\alpha,\alpha_0\}) = 0, \qquad (2.3.36)$$

we only need to consider $X_i \in [0,\alpha_m(\alpha)) \cup (\alpha_M(\alpha),1]$. Note that for all $X_i \notin (\alpha_m(\alpha),\alpha_M(\alpha))$ we have

$$\log(f(X_i,\alpha,\theta)) - \log(f(X_i,\alpha_0,\theta)) = \begin{cases} \log\beta(\alpha,\theta) - \log\beta(\alpha_0,\theta), & \text{if } X_i < \alpha_m(\alpha), \\ \log\gamma(\alpha,\theta) - \log\gamma(\alpha_0,\theta), & \text{if } X_i > \alpha_M(\alpha) \end{cases} \qquad (2.3.37)$$

$$= \begin{cases} \log\frac{\alpha_0}{\alpha}, & \text{if } X_i < \alpha_m(\alpha), \\ \log\frac{1-\alpha_0}{1-\alpha}, & \text{if } X_i > \alpha_M(\alpha). \end{cases} \qquad (2.3.38)$$

Notice that this value does not depend on θ. We can linearise this with respect to α around the point $\alpha = \alpha_0$ as for all X_i that we consider, Equation (2.3.38) is differentiable with respect to α around a neighbourhood of α_0. Define

$$b(x, \alpha) := \begin{cases} \log \frac{\alpha_0}{\alpha}, & \text{if } x < \alpha_m(\alpha), \\ \log \frac{1-\alpha_0}{1-\alpha}, & \text{if } x > \alpha_M(\alpha). \end{cases} \tag{2.3.39}$$

Then linearization around α_0 yields

$$b(x, \alpha) = b(x, \alpha_0) + \frac{\partial b(x, \alpha_0)}{\partial \alpha}(\alpha - \alpha_0) + h(x, \alpha)(\alpha - \alpha_0), \tag{2.3.40}$$

for some function $h(x, \alpha)$. We can clearly see that for any x we have

$$b(x, \alpha_0) = 0. \tag{2.3.41}$$

Therefore

$$h(x, \alpha) = \frac{b(x, \alpha)}{\alpha - \alpha_0} - \frac{\partial b(x, \alpha_0)}{\partial \alpha}. \tag{2.3.42}$$

The partial derivative of b with respect to α is given by

$$\frac{\partial b(x, \alpha)}{\partial \alpha} = \begin{cases} -\frac{1}{\alpha}, & \text{if } x < \alpha_m(\alpha), \\ \frac{1}{1-\alpha}, & \text{if } x > \alpha_M(\alpha). \end{cases} \tag{2.3.43}$$

So for $x < \alpha_m(\alpha)$ we have

$$h(x, \alpha) = h_1(\alpha) := \frac{\log \frac{\alpha_0}{\alpha}}{\alpha - \alpha_0} + \frac{1}{\alpha_0}, \tag{2.3.44}$$

and for $x > \alpha_M(\alpha)$ we have

$$h(x, \alpha) = h_2(\alpha) := \frac{\log \frac{1-\alpha_0}{1-\alpha}}{\alpha - \alpha_0} - \frac{1}{1 - \alpha_0}. \tag{2.3.45}$$

By using l'Hôpital's rule we see that

$$\lim_{\alpha \to \alpha_0} h_1(\alpha) = \lim_{\alpha \to \alpha_0} -\frac{1}{\alpha} + \frac{1}{\alpha_0} = 0, \tag{2.3.46}$$

and

$$\lim_{\alpha \to \alpha_0} h_2(\alpha) = \lim_{\alpha \to \alpha_0} \frac{1}{1 - \alpha} - \frac{1}{1 - \alpha_0} = 0. \tag{2.3.47}$$

Hence both h_1 and h_2 are $o(1)$ as $\alpha \to \alpha_0$. Using b in the definition of $S_n^{(2)}(\alpha,\theta)$, we find that

$$S_n^{(2)}(\alpha,\theta) = \frac{1}{n} \sum_{X_i \notin (\alpha_m(\alpha), \alpha_M(\alpha))} b(X_i,\alpha) \tag{2.3.48}$$

$$= \frac{1}{n} \sum_{X_i \notin (\alpha_m(\alpha), \alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) + \frac{1}{n}\sum_{X_i<\alpha_m} h_1(\alpha)(\alpha-\alpha_0) + \frac{1}{n}\sum_{X_i>\alpha_M} h_2(\alpha)(\alpha-\alpha_0) \tag{2.3.49}$$

$$= \frac{1}{n} \sum_{X_i \notin (\alpha_m(\alpha), \alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) + \big(F_n(\alpha_m(\alpha))h_1(\alpha) + (1-F_n(\alpha_M(\alpha)))h_2(\alpha)\big)(\alpha-\alpha_0). \tag{2.3.50}$$

Recall the definition of T_p from Definition 2.1. We now want to show that

$$\big(F_n(\alpha_m(\alpha))h_1(\alpha) + (1-F_n(\alpha_M(\alpha)))h_2(\alpha)\big)(\alpha-\alpha_0) = T_p(|\alpha-\alpha_0|), \tag{2.3.51}$$

for $\alpha \to \alpha_0$. To do this we assume η_n is some consistent estimator for α_0. We then analyse

$$\big|F_n(\alpha_m(\eta_n))h_1(\eta_n) + (1-F_n(\alpha_M(\eta_n)))h_2(\eta_n)\big|. \tag{2.3.52}$$

For any $n \in \mathbb{N}$ and $\alpha \in (0,1)$ we know that by definition of a CDF $|F_n(\alpha_m(\alpha))|$ and $|1-F_n(\alpha_M(\alpha))|$ are bounded by 1. So these sequences are also bounded by 1 for $\alpha = \eta_n$. Hence

$$\big|F_n(\alpha_m(\eta_n))h_1(\eta_n) + (1-F_n(\alpha_M(\eta_n)))h_2(\eta_n)\big| \leq |h_1(\eta_n)| + |h_2(\eta_n)|. \tag{2.3.53}$$

Since $h_1(\alpha)$ and $h_2(\alpha)$ are both $o(1)$ as $\alpha \to \alpha_0$ and do not depend on n, we can use consistency of η_n to find that that $h_1(\eta_n)$ and $h_2(\eta_n)$ are $o_p(1)$ as $n \to \infty$. Hence

$$F_n(\alpha_m(\eta_n))h_1(\eta_n) + (1-F_n(\alpha_M(\eta_n)))h_2(\eta_n) = o_p(1), \tag{2.3.54}$$

for $n \to \infty$. Therefore indeed

$$\big(F_n(\alpha_m(\alpha))h_1(\alpha) + (1-F_n(\alpha_M(\alpha)))h_2(\alpha)\big)(\alpha-\alpha_0) = T_p(|\alpha-\alpha_0|), \tag{2.3.55}$$

for $\alpha \to \alpha_0$. Putting this back into the expression for $S_n^{(2)}(\alpha,\theta)$ that we found in Equation (2.3.50) yields

$$S_n^{(2)}(\alpha,\theta) = \frac{1}{n} \sum_{X_i \notin (\alpha_m(\alpha), \alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) + T_p(|\alpha-\alpha_0|), \tag{2.3.56}$$

for $\alpha \to \alpha_0$. Now we want to rewrite the sum that we see in the equation above. To do this we rewrite it as the sum over all observations and subtract the observations between α and α_0. This gives us

$$\begin{aligned}\frac{1}{n}\sum_{X_i \notin (\alpha_m(\alpha),\alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) \\ &\quad - \frac{1}{n}\sum_{X_i \in (\alpha_m(\alpha),\alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0).\end{aligned} \tag{2.3.57}$$

We want to show that

$$\frac{1}{n}\sum_{X_i \in (\alpha_m(\alpha),\alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) = T_p(|\alpha-\alpha_0|), \tag{2.3.58}$$

for $\alpha \to \alpha_0$. By Equation (2.3.43), we have

$$\frac{1}{n}\sum_{X_i \in (\alpha_m(\alpha),\alpha_M(\alpha))} \frac{\partial b(X_i,\alpha_0)}{\partial \alpha}(\alpha-\alpha_0) = \begin{cases} \frac{\alpha-\alpha_0}{1-\alpha_0}(F_n(\alpha)-F_n(\alpha_0)), & \text{if } \alpha > \alpha_0, \\ \frac{\alpha-\alpha_0}{\alpha_0}(F_n(\alpha)-F_n(\alpha_0)), & \text{if } \alpha < \alpha_0. \end{cases} \tag{2.3.59}$$

Dividing by α_0 or $1-\alpha_0$ does not have any effects on the asymptotic properties. Therefore it suffices to show that

$$(\alpha-\alpha_0)(F_n(\alpha)-F_n(\alpha_0)) = T_p(|\alpha-\alpha_0|), \tag{2.3.60}$$

for $\alpha \to \alpha_0$. Clearly we can instead show that

$$F_n(\alpha)-F_n(\alpha_0) = T_p(1), \tag{2.3.61}$$

for $\alpha \to \alpha_0$. To do this we let η_n be a consistent estimator of α_0. Then

$$F_n(\eta_n)-F_n(\alpha_0) \sim \frac{1}{n}\text{Binom}(n,|\eta_n-\alpha_0|). \tag{2.3.62}$$

Let $\epsilon > 0$. By consistency of η_n, we can choose $N \in \mathbb{N}$ such that for all $n \geq N$ we have $|\eta_n - \alpha_0| < \epsilon/2$. Then for $n \geq N$

$$\mathbb{P}(F_n(\eta_n)-F_n(\alpha_0) < \epsilon) \geq \mathbb{P}\left(\frac{1}{n}\text{Binom}(n,\epsilon/2) < \epsilon\right). \tag{2.3.63}$$

By the law of large numbers

$$\frac{1}{n}\text{Binom}(n,\epsilon/2) \xrightarrow{a.s.} \epsilon/2. \tag{2.3.64}$$

Therefore

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\text{Binom}(n,\epsilon/2) < \epsilon\right) = 1. \tag{2.3.65}$$

By the squeeze theorem we thus conclude that

$$\lim_{n\to\infty} \mathbb{P}(F_n(\eta_n) - F_n(\alpha_0) < \epsilon) = 1, \tag{2.3.66}$$

and hence

$$F_n(\alpha) - F_n(\alpha_0) = T_p(1), \tag{2.3.67}$$

and

$$\frac{1}{n} \sum_{X_i \in (\alpha_m(\alpha), \alpha_M(\alpha))} \frac{\partial b(X_i, \alpha_0)}{\partial \alpha}(\alpha - \alpha_0) = T_p(|\alpha - \alpha_0|), \tag{2.3.68}$$

for $\alpha \to \alpha_0$. Therefore

$$\begin{aligned} S_n^{(2)}(\alpha, \theta) &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial b(X_i, \alpha_0)}{\partial \alpha}(\alpha - \alpha_0) + T_p(|\alpha - \alpha_0|) + T_p(|\alpha - \alpha_0|) && (2.3.69) \\ &= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial b(X_i, \alpha_0)}{\partial \alpha}(\alpha - \alpha_0) + T_p(|\alpha - \alpha_0|), && (2.3.70) \end{aligned}$$

for $\alpha \to \alpha_0$. By the law of large numbers we have

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial b(X_i, \alpha_0)}{\partial \alpha} \xrightarrow{\mathbb{P}} \mathbb{E}\left[\frac{\partial b(X, \alpha_0)}{\partial \alpha}\right], \tag{2.3.71}$$

where X is a random variable which is distributed with PDF $f(x, \alpha_0, \theta_0)$. This can also be written as

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial b(X_i, \alpha_0)}{\partial \alpha} = \mathbb{E}\left[\frac{\partial b(X, \alpha_0)}{\partial \alpha}\right] + o_p(1), \tag{2.3.72}$$

as $n \to \infty$. Since

$$\begin{aligned} \mathbb{E}\left[\frac{\partial b(X, \alpha_0)}{\partial \alpha}\right] &= \mathbb{P}(X < \alpha_0) \cdot \frac{-1}{\alpha_0} + \mathbb{P}(X > \alpha_0) \cdot \frac{1}{1 - \alpha_0} && (2.3.73) \\ &= \frac{1 - \theta_0}{1 - \alpha_0} - \frac{\theta_0}{\alpha_0} && (2.3.74) \\ &= \gamma_0 - \beta_0, && (2.3.75) \end{aligned}$$

we find that

$$S_n^{(2)}(\alpha, \theta) = (\alpha - \alpha_0)(\gamma_0 - \beta_0) + (\alpha - \alpha_0) o_p(1) + T_p(|\alpha - \alpha_0|), \tag{2.3.76}$$

for $n \to \infty$.

Now we want to combine $S_n^{(1)}(\alpha, \theta)$ and $S_n^{(2)}(\alpha, \theta)$ to get an expression for $S_n(\alpha, \theta)$. We find that

$$\begin{aligned} S_n(\alpha, \theta) = (F_n(\alpha) - F_n(\alpha_0)) \log\left(\frac{\beta(\alpha_0, \theta)}{\gamma(\alpha_0, \theta)}\right) + o(1) M_n(\alpha, \theta) + (\alpha - \alpha_0) o(1) \\ + (\alpha - \alpha_0)(\gamma_0 - \beta_0) + (\alpha - \alpha_0) o_p(1) + T_p(|\alpha - \alpha_0|), \end{aligned} \tag{2.3.77}$$

where o and T_p are for $\alpha \to \alpha_0$ and o_p is for $n \to \infty$. Since

$$(\alpha-\alpha_0)(\gamma_0-\beta_0)+(F_n(\alpha)-F_n(\alpha_0))\log\left(\frac{\beta(\alpha_0,\theta)}{\gamma(\alpha_0,\theta)}\right)=\log\left(\frac{\beta(\alpha_0,\theta)}{\gamma(\alpha_0,\theta)}\right)M_n(\alpha,\theta), \quad (2.3.78)$$

we can conclude that

$$\begin{aligned} S_n(\alpha,\theta) &= (\log(\beta(\alpha_0,\theta))-\log(\gamma(\alpha_0,\theta))+o(1))M_n(\alpha,\theta) \\ &\quad + (\alpha-\alpha_0)(o(1)+o_p(1))+T_p(|\alpha-\alpha_0|), \end{aligned} \quad (2.3.79)$$

where o and T_p are for $\alpha \to \alpha_0$ and o_p is for $n \to \infty$. ■

2.4 Convergence speed of QMLE

Recall that the QMLE is defined in Definition 2.2 as

$$\tilde{\alpha}_n := \underset{\alpha\in[\alpha_0-a_n,\alpha_0+a_n]}{\arg\max} M_n(\alpha,\theta_0), \quad (2.4.1)$$

where the function $M_n(\alpha,\theta)$ is defined in Equation (2.3.1) as

$$M_n(\alpha,\theta) := F_n(\alpha)-F_n(\alpha_0)+\frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta))-\log(\gamma(\alpha_0,\theta))}. \quad (2.4.2)$$

The goal of this section is to prove the following theorem.

Theorem 2.9. *The QMLE has the property that*

$$\tilde{\alpha}_n-\alpha_0=O_p\left(\frac{1}{n}\right), \quad (2.4.3)$$

as $n \to \infty$.

First we will give a heuristic overview of the proof while introducing a relevant lemma. After that we will give the detailed proof of the theorem. The lemma will be proven afterwards.

2.4.1 Proof of Theorem 2.9

The proof of Theorem 2.9 is based on the following lemma, which will be proven in the next subsection.

Lemma 2.10. *For each* $\epsilon > 0$ *there exist* $K, k > 0$ *and* $N \in \mathbb{N}$ *such that for all* $n \geq N$ *we have*

$$\mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|<a_n}\frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|}<-k\right)>1-\epsilon. \quad (2.4.4)$$

Lemma 2.10 gives us a stronger statement than needed in the proof of the theorem. We do state it in this form, because this lemma will be used again in Section 2.5. In this case we use that this lemma roughly says that $\sup_{K/n<|\alpha-\alpha_0|<a_n} M_n(\alpha,\theta_0) < 0$ eventually for some K. Since $M_n(\alpha_0,\theta_0) = 0$, the definition of $\tilde{\alpha}_n$ tells us that $M_n(\tilde{\alpha}_n,\theta_0) \geq 0$. Hence $n|\tilde{\alpha}_n - \alpha_0| < K$.

Proof of Theorem 2.9. We start by reducing the strong statement from Lemma 2.10 into a statement that will be mmore useful to the proof. Lemma 2.10 states that for each $\epsilon > 0$ there exist $K, k > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k\right) > 1-\epsilon. \tag{2.4.5}$$

If for some $K, k > 0$ and $n \in \mathbb{N}$ we have that

$$\sup_{K/n<|\alpha-\alpha_0|<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k, \tag{2.4.6}$$

then we also have

$$\sup_{K/n<|\alpha-\alpha_0|<a_n} M_n(\alpha,\theta_0) < 0. \tag{2.4.7}$$

Thus by Lemma 2.10 we find that for all $\epsilon > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|<a_n} M_n(\alpha,\theta_0) < 0\right) > 1-\epsilon. \tag{2.4.8}$$

Now we can use this statement to prove the final result.

Let $\epsilon > 0$. Choose $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ Equation (2.4.8) holds. Notice that

$$M_n(\alpha_0,\theta_0) = 0. \tag{2.4.9}$$

Therefore using that by definition $\tilde{\alpha}_n$ maximises $M_n(\alpha,\theta_0)$ on $[\alpha_0 - a_n, \alpha_0 + a_n]$, we see that

$$M_n(\tilde{\alpha}_n,\theta_0) \geq M_n(\alpha_0,\theta_0) = 0. \tag{2.4.10}$$

By definition, we always have $|\tilde{\alpha}_n - \alpha_0| < a_n$. Hence for any $n \in \mathbb{N}$

$$n|\tilde{\alpha}_n - \alpha_0| > K \implies \sup_{K/n<|\alpha-\alpha_0|<a_n} M_n(\alpha,\theta_0) \geq 0. \tag{2.4.11}$$

Thus for all $n \geq N$ we have

$$\mathbb{P}(n|\tilde{\alpha}_n - \alpha_0| \geq K) \leq \mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|<a_n} M_n(\alpha,\theta_0) \geq 0\right) < \epsilon. \tag{2.4.12}$$

We conclude that indeed

$$\tilde{\alpha}_n - \alpha_0 = O_p\left(\frac{1}{n}\right), \tag{2.4.13}$$

as $n \to \infty$. ∎

2.4.2 Proof of Lemma 2.10

We now continue by proving Lemma 2.10. This lemma states that for each $\epsilon > 0$ there exist $K, k > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k \right) > 1 - \epsilon. \tag{2.4.14}$$

First we give a heuristic overview of the proof and introduce a relevant lemma. Then we will give the detailed proof. The lemma that we introduce will be proven in the next subsection.

We split the supremum into two parts. The first part considers $\alpha > \alpha_0$ and the second part considers $\alpha < \alpha_0$. If we can bound both of these suprema, we have bound the entire supremum. The two cases can are analogous to each other. We will cover the case of $\alpha > \alpha_0$. First we use the definition of M_n and apply some computational steps to find that

$$\frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} = \left(\frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right) \gamma_0 + \gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0}. \tag{2.4.15}$$

It can be proven that

$$\gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} < 0. \tag{2.4.16}$$

We then use the following theorem, which will be proven in the next subsection.

Lemma 2.11. *For all $\epsilon, k > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have*

$$\mathbb{P}\left(\sup_{K/n < |\alpha - \alpha_0|} \left| \frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right| < k \right) > 1 - \epsilon. \tag{2.4.17}$$

We apply Lemma 2.11 by choosing

$$k = -\frac{1}{2\gamma_0} \left(\gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \right) > 0. \tag{2.4.18}$$

Then we find that

$$\frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < \frac{1}{2} \left(\gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \right) < 0. \tag{2.4.19}$$

Analogously we can find a bound for $\alpha < \alpha_0$. Thus the whole supremum is bounded.

Now we will give the detailed proof.

Proof of Lemma 2.10. We will look at $\alpha > \alpha_0$ and $\alpha < \alpha_0$ separately. These correspond to

$$\mathbb{P}\left(\sup_{K/n < \alpha - \alpha_0 < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k \right), \tag{2.4.20}$$

and

$$\mathbb{P}\left(\sup_{K/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k \right), \tag{2.4.21}$$

respectively. We will find a different K, k, and N for both probabilities to be less than $1-\epsilon/2$. The two cases are analogous. To find these K, k, and N, we will write $M_n(\alpha,\theta_0)/|\alpha-\alpha_0|$ as a sum of a negative constant and an expression that, with great probability, is less than minus the negative constant for some K and N. The fact that this second part of the sum has this property will be proven by Lemma 2.11.

Fix $\epsilon > 0$. We will first look at $\alpha > \alpha_0$. Notice that for $\alpha > \alpha_0$ we have

$$\begin{aligned}
\frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} &= \frac{F_n(\alpha)-F_n(\alpha_0)}{\alpha-\alpha_0} + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} && (2.4.22)\\
&= \frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} \cdot \frac{F_0(\alpha)-F_0(\alpha_0)}{\alpha-\alpha_0} + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} && (2.4.23)\\
&= \frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} \cdot \frac{\gamma_0(\alpha-\alpha_0)}{\alpha-\alpha_0} + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} && (2.4.24)\\
&= \frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}\gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} && (2.4.25)\\
&= \left(\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right)\gamma_0 + \gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0}. && (2.4.26)
\end{aligned}$$

We will to prove that

$$\gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} < 0. \tag{2.4.27}$$

Observe that

$$\begin{aligned}
\gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} &= \frac{1-\alpha_0}{1-\theta_0} + \frac{\frac{1-\alpha_0}{1-\theta_0} - \frac{\alpha_0}{\theta_0}}{\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)}} && (2.4.28)\\
&= \frac{1}{\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)}}\left(\frac{1-\alpha_0}{1-\theta_0}\cdot\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} + \frac{1-\alpha_0}{1-\theta_0} - \frac{\alpha_0}{\theta_0}\right) && (2.4.29)\\
&= \frac{1-\alpha_0}{(1-\theta_0)\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)}}\left(\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} + 1 - \frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)}\right). && (2.4.30)
\end{aligned}$$

Since $\beta_0 > \gamma_0$ by Assumption 2.2, we have $\alpha_0 > \theta_0$ and

$$\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} > 1. \tag{2.4.31}$$

Therefore

$$\frac{1-\alpha_0}{(1-\theta_0)\log\frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)}} > 0. \tag{2.4.32}$$

Hence we only need to prove that

$$\log \frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} + 1 - \frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} < 0. \tag{2.4.33}$$

Define $g : (0, \infty) \to \mathbb{R}$ by

$$g(x) := 1 - x + \log x. \tag{2.4.34}$$

Clearly $g(1) = 0$ and $g'(x) = \frac{1}{x} - 1$. For $x > 1$, we thus have $g'(x) < 0$. Hence for $x > 1$ we get $g(x) < 0$. Hence using Equation (2.4.31) we find that

$$\log \frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} + 1 - \frac{\alpha_0(1-\theta_0)}{\theta_0(1-\alpha_0)} < 0. \tag{2.4.35}$$

Therefore

$$\gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} < 0. \tag{2.4.36}$$

Now we want to bound the first part of Equation (2.4.26), which is given by

$$\left(\frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right) \gamma_0. \tag{2.4.37}$$

To do this, we shall use Lemma 2.11, which states that for all $\epsilon', k > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{K/n < |\alpha - \alpha_0|} \left| \frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right| < k \right) > 1 - \epsilon'. \tag{2.4.38}$$

Let

$$k_1 := -\frac{1}{2} \left(\gamma_0 + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \right) > 0. \tag{2.4.39}$$

By the hypothesis in equation Lemma 2.11 there exist $K_1 > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{K_1/n < |\alpha - \alpha_0|} \left| \frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right| < \frac{k_1}{\gamma_0} \right) > 1 - \frac{\epsilon}{2}. \tag{2.4.40}$$

Thus for all $n \geq N_1$ we also have

$$\mathbb{P}\left(\sup_{K_1/n < \alpha - \alpha_0} \left| \frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right| < \frac{k_1}{\gamma_0} \right) > 1 - \frac{\epsilon}{2}. \tag{2.4.41}$$

If for some $n \in \mathbb{N}$

$$\left| \frac{F_n(\alpha) - F_n(\alpha_0)}{F_0(\alpha) - F_0(\alpha_0)} - 1 \right| < \frac{k_1}{\gamma_0}, \tag{2.4.42}$$

then

$$\frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} = \left(\frac{F_n(x)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right)\gamma_0 + \gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} \tag{2.4.43}$$

$$\leq \left|\frac{F_n(x)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right|\gamma_0 + \gamma_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} \tag{2.4.44}$$

$$< -k_1. \tag{2.4.45}$$

Therefore for all $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_1\right) \geq \mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0} \left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right| < \frac{k_1}{\gamma_0}\right) \tag{2.4.46}$$

$$> 1 - \frac{\epsilon}{2}. \tag{2.4.47}$$

We continue by looking at the case $\alpha < \alpha_0$. Using a completely similar argumentation we can show that there exist $K_2, k_2 > 0$ and $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$ we have

$$\mathbb{P}\left(\sup_{K_2/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_2\right) > 1 - \frac{\epsilon}{2}. \tag{2.4.48}$$

There is only a slight difference. This time we notice that for $\alpha < \alpha_0$

$$M_n(\alpha,\theta_0) = \left(1 - \frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}\right)\beta_0 - \left(\beta_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0}\right), \tag{2.4.49}$$

and prove that

$$\beta_0 + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0} > 0. \tag{2.4.50}$$

Since this proof is analogous to the previous case, we do not repeat the proof. Now we can start to combine the results for the two cases in order to get to the final conclusion.

Define $K_0 := \max\{K_1, K_2\}$, $k_0 := \min\{k_1, k_2\}$, and $N_0 := \max\{N_1, N_2\}$. Then

$$\sup_{K_1/n<\alpha-\alpha_0<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_1 \implies \sup_{K_0/n<\alpha-\alpha_0<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_0, \tag{2.4.51}$$

and

$$\sup_{K_2/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_2 \implies \sup_{K_0/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_0. \tag{2.4.52}$$

Hence if

$$\sup_{K_1/n<\alpha-\alpha_0<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_1, \quad \text{and} \quad \sup_{K_2/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_2, \tag{2.4.53}$$

then

$$\sup_{K_0/n<|\alpha-\alpha_0|<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_0. \tag{2.4.54}$$

We now use the Fréchet inequalities and Equations (2.4.47) and (2.4.48) to find that for $n \geq N_0$

$$\begin{aligned}
&\mathbb{P}\left(\sup_{K_0/n<|\alpha-\alpha_0|<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_0\right) \\
&\quad \geq \mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_1, \quad \sup_{K_2/n<\alpha_0-\alpha<a_n} \frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|} < -k_2\right) \\
&\quad \geq 1-\frac{\epsilon}{2}+1-\frac{\epsilon}{2}-1 \\
&\quad = 1-\epsilon.
\end{aligned}$$

■

2.4.3 Proof of Lemma 2.11

Finally we will prove Lemma 2.11. Recall that this lemma states that for all $\epsilon, k > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|} \left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right| < k\right) > 1-\epsilon. \tag{2.4.55}$$

To prove this we start by giving a short outline of the proof an introducing a lemma that we will use. This lemma is proven in the appendix. Then we will give the proof of Lemma 2.11 in full detail.

Just like in the proof for the previous lemma look at $\alpha < \alpha_0$ and $\alpha > \alpha_0$ separately. The two cases are completely analogous. We first consider $\alpha > \alpha_0$. We find that

$$\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right| = \left|\frac{F_n(\alpha)-F_n(\alpha_0)}{\gamma_0(\alpha-\alpha_0)} - 1\right|. \tag{2.4.56}$$

However, the number of points in an interval that lies completely on one side of α_0 is distributed according to a binomial distribution. The number of observations in any interval from a uniform distribution is distributed according to a binomial distribution. This observation shows us that for any $K, k > 0$ and $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{K/n<\alpha-\alpha_0} \left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)} - 1\right| < k\right) = \mathbb{P}\left(\sup_{\frac{K}{n\gamma_0}<x} \left|\frac{G_n(x)}{x} - 1\right| < k\right). \tag{2.4.57}$$

At this point we use the following lemma, which is proven in Appendix Section A.2.

Lemma 2.12. *For all $\epsilon, k > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have*

$$\mathbb{P}\left(\sup_{x>K/n}\left|\frac{G_n(x)}{x}-1\right|<k\right)>1-\epsilon, \tag{2.4.58}$$

where G_n is the eCDF for the uniform distribution on $[0,1]$.

This bound on the supremum of $\left|\frac{G_n(x)}{x}-1\right|$ directly gives us a bound on the supremum of $\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|$. Repeating this for $\alpha < \alpha_0$ gives us the final result. Now we move on to the detailed proof.

Proof of Lemma 2.11. Let $\epsilon, k > 0$ be fixed. First we consider $\alpha > \alpha_0$. Notice that

$$F_0(\alpha)-F_0(\alpha_0)=1-\gamma_0(1-\alpha)-1+\gamma_0(1-\alpha_0)=\gamma_0(\alpha-\alpha_0). \tag{2.4.59}$$

Thus

$$\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|=\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{\gamma_0(\alpha-\alpha_0)}-1\right|. \tag{2.4.60}$$

Notice that $F_n(\alpha)-F_n(\alpha_0)$ is the proportion of the observations that lie in $(\alpha_0, \alpha]$. The probability of any given sample to fall in that interval is $\gamma_0(\alpha-\alpha_0)$. Therefore

$$F_n(\alpha)-F_n(\alpha_0)\sim \text{Binom}(n,\gamma_0(\alpha-\alpha_0)). \tag{2.4.61}$$

Define G_n to be the eCDF of some uniform distribution on $[0,1]$. Then we also have

$$G_n(\gamma_0(\alpha-\alpha_0))\sim \text{Binom}(n,\gamma_0(\alpha-\alpha_0)). \tag{2.4.62}$$

Hence for any $K_1 > 0$ and $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k\right)=\mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{G_n(\gamma_0(\alpha-\alpha_0))}{\gamma_0(\alpha-\alpha_0)}-1\right|<k\right). \tag{2.4.63}$$

By taking $x=\gamma_0(\alpha-\alpha_0)$ we find that for any $K_1 > 0$ and $n \in \mathbb{N}$

$$\mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{G_n(\gamma_0(\alpha-\alpha_0))}{\gamma_0(\alpha-\alpha_0)}-1\right|<k\right)=\mathbb{P}\left(\sup_{\frac{K_1}{n\gamma_0}<x}\left|\frac{G_n(x)}{x}-1\right|<k\right). \tag{2.4.64}$$

Now we can use Lemma 2.12. By this lemma, we can take some $K_1' > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$

$$\mathbb{P}\left(\sup_{\frac{K_1'}{n}<x}\left|\frac{G_n(x)}{x}-1\right|<k\right)>1-\frac{\epsilon}{2}. \tag{2.4.65}$$

By taking $K_1 = \gamma_0 K_1'$ we get that for all $n \geq N_1$

$$\mathbb{P}\left(\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k\right)>1-\frac{\epsilon}{2}. \tag{2.4.66}$$

We can now consider $\alpha < \alpha_0$. In a completely analogous fashion it can be shown that there exist $K_2 > 0$ and $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$ we have

$$\mathbb{P}\left(\sup_{K_2/n<\alpha_0-\alpha}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k\right)>1-\frac{\epsilon}{2}. \tag{2.4.67}$$

Now we can combine the two cases to get to the final conclusion.

Define $K_0 := \max\{K_1, K_2\}$ and $N_0 := \max\{N_1, N_2\}$. Clearly

$$\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k \implies \sup_{K_0/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k, \tag{2.4.68}$$

and

$$\sup_{K_2/n<\alpha_0-\alpha}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k \implies \sup_{K_0/n<\alpha_0-\alpha}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k. \tag{2.4.69}$$

Hence if

$$\sup_{K_1/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k \quad \text{and} \quad \sup_{K_2/n<\alpha_0-\alpha}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k, \tag{2.4.70}$$

then

$$\sup_{K_0/n<|\alpha-\alpha_0|}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k. \tag{2.4.71}$$

Thus we can use the Fréchet inequalities to find that for $n \geq N_0$

$$\mathbb{P}\left(\sup_{K_0/n<|\alpha-\alpha_0|}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k\right) \tag{2.4.72}$$

$$\geq \mathbb{P}\left(\sup_{K_0/n<\alpha-\alpha_0}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k, \quad \sup_{K_0/n<\alpha_0-\alpha}\left|\frac{F_n(\alpha)-F_n(\alpha_0)}{F_0(\alpha)-F_0(\alpha_0)}-1\right|<k\right) \tag{2.4.73}$$

$$\geq 1-\epsilon. \tag{2.4.74}$$

■

2.5 Convergence speed of MLE

The goal of this section is to show the following theorem.

Theorem 2.13. *The MLE for α_0 has the property that*

$$\hat{\alpha}_n = \alpha_0 + O_p\left(\frac{1}{n}\right), \tag{2.5.1}$$

as $n \to \infty$.

The proof of this theorem will combine Theorem 2.9, which says that

$$\tilde{\alpha}_n = \alpha_0 + O_p\left(\frac{1}{n}\right), \tag{2.5.2}$$

with the following theorem.

Theorem 2.14. *The MLE and QMLE for α_0 have the property that*

$$\hat{\alpha}_n - \tilde{\alpha}_n = o_p(1/n). \tag{2.5.3}$$

The proof of Theorem 2.13 is relatively short, so we start by stating this proof. Afterwards we will prove Theorem 2.14.

Proof of Theorem 2.13. By Theorem 2.14, we know that

$$\hat{\alpha}_n = \tilde{\alpha}_n + o_p\left(\frac{1}{n}\right), \tag{2.5.4}$$

as $n \to \infty$. However by Theorem 2.9

$$\tilde{\alpha}_n = \alpha_0 + O_p\left(\frac{1}{n}\right), \tag{2.5.5}$$

as $n \to \infty$. Hence

$$\hat{\alpha}_n = \alpha_0 + O_p\left(\frac{1}{n}\right) + o_p\left(\frac{1}{n}\right) = \alpha_0 + O_p\left(\frac{1}{n}\right), \tag{2.5.6}$$

as $n \to \infty$. ■

2.5.1 Preparations for proving Theorem 2.14

In order to prove Theorem 2.14, we will look at the difference of the log-likelihood evaluated in $(\hat{\alpha}_n, \hat{\theta}_n)$ and the log-likelihood evaluated in $(\tilde{\alpha}_n, \hat{\theta}_n)$. We will rewrite this difference to a usable expression by using some small lemmas that we will prove along the way. Then we move on to the next subsection where we use this result in the proof of Theorem 2.14.

We start by considering the log-likelihood. Recall from Theorem 2.7 that the log-likelihood is given by

$$\ell_n(\alpha, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log(f(X_i, \alpha_0, \theta)) + S_n(\alpha, \theta), \tag{2.5.7}$$

where

$$\begin{aligned} S_n(\alpha, \theta) = (\log(\beta(\alpha_0, \theta)) - \log(\gamma(\alpha_0, \theta)) + o(1)) M_n(\alpha, \theta) \\ + (\alpha - \alpha_0)(o(1) + o_p(1)) + T_p(|\alpha - \alpha_0|), \end{aligned} \tag{2.5.8}$$

where o and T_p are for $\alpha \to \alpha_0$ and o_p is for $n \to \infty$. Since by definition the MLEs $\hat{\alpha}_n$ and $\hat{\theta}_n$ maximise the log-likelihood,

$$\ell_n(\hat{\alpha}_n, \hat{\theta}_n) - \ell_n(\tilde{\alpha}_n, \hat{\theta}_n) \geq 0. \tag{2.5.9}$$

However,

$$\ell_n(\hat{\alpha}_n, \hat{\theta}_n) - \ell_n(\tilde{\alpha}_n, \hat{\theta}_n) = S_n(\hat{\alpha}_n, \hat{\theta}_n) - S_n(\tilde{\alpha}_n, \hat{\theta}_n). \tag{2.5.10}$$

Therefore

$$0 \leq S_n(\hat{\alpha}_n, \hat{\theta}_n) - S_n(\tilde{\alpha}_n, \hat{\theta}_n). \tag{2.5.11}$$

Recall that by the definition of T_p evaluating $T_p(|\alpha - \alpha_0|)$ in a consistent estimator η_n for α_0 yields a function that is $o_p(|\eta_n - \alpha_0|)$. By Theorem 2.2 we know that $\hat{\alpha}$ is consistent. Recall that

$$\tilde{\alpha}_n = \underset{\alpha \in [\alpha_0 - a_n, \alpha_0 + a_n]}{\arg\max} M_n(\alpha, \theta_0), \tag{2.5.12}$$

and $\lim_{n\to\infty} a_n = 0$. Therefore $\tilde{\alpha}_n$ is also consistent. Thus we know what happens when we evaluate $T_p(|\alpha - \alpha_0|)$ in $\hat{\alpha}_n$ or $\tilde{\alpha}_n$.

We still need to look at what happens when we evaluate functions that are $o(1)$ as $\alpha \to \alpha_0$ in $\hat{\alpha}_n$ or $\tilde{\alpha}_n$. It will give us some asymptotic behaviour in probability as $n \to \infty$. For this we use the following lemma.

Lemma 2.15. *Suppose that $h \in o(1)$ for $\alpha \to \alpha_0$ and η_n is a consistent estimator of α_0. Then*

$$h(\eta_n) = o_p(1), \tag{2.5.13}$$

for $n \to \infty$.

Proof. Let $\epsilon_1, \epsilon_2 > 0$. Since $h \in o(1)$, there exists $\delta > 0$ such that for all α with $|\alpha - \alpha_0| < \delta$ we have $|h(\alpha)| < \epsilon_1$. Since $\eta_n - \alpha_0 = o_p(1)$ as $n \to \infty$, there exists $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\mathbb{P}(|\eta_n - \alpha_0| > \delta) < \epsilon_2. \tag{2.5.14}$$

Since

$$|\eta_n - \alpha_0| < \delta \implies |h(\eta_n)| < \epsilon_1, \tag{2.5.15}$$

we have

$$|h(\eta_n)| > \epsilon_1 \implies |\eta_n - \alpha_0| > \delta. \tag{2.5.16}$$

Therefore for $n > N$ we have

$$\mathbb{P}(|h(\eta_n)| > \epsilon_1) \leq \mathbb{P}(|\eta_n - \alpha_0| > \delta) < \epsilon_2. \tag{2.5.17}$$

Hence

$$h(\eta_n) = o_p(1). \tag{2.5.18}$$

■

Since $\hat{\alpha}_n$ and $\tilde{\alpha}_n$ are consistent estimators of α_0, Lemma 2.15 tell us that evaluating a function that is $o(1)$ for $\alpha \to \alpha_0$ in either $\hat{\alpha}_n$ or $\tilde{\alpha}_n$ yields a function that is $o_p(1)$ for $n \to \infty$.

Therefore

$$\begin{aligned} S_n(\hat{\alpha}_n, \hat{\theta}_n) - S_n(\tilde{\alpha}_n, \hat{\theta}_n) = {} & (\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)) + o_p(1))(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) \\ & + (\hat{\alpha}_n - \tilde{\alpha}_n)o_p(1) + o_p(|\hat{\alpha}_n - \alpha_0|) + o_p(|\tilde{\alpha}_n - \alpha_0|), \end{aligned} \tag{2.5.19}$$

for $n \to \infty$. Since $0 \leq S_n(\hat{\alpha}_n, \hat{\theta}_n) - S_n(\tilde{\alpha}_n, \hat{\theta}_n)$, we find that

$$\begin{aligned} 0 \leq {} & (\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)) + o_p(1))(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) \\ & + (\hat{\alpha}_n - \tilde{\alpha}_n)o_p(1) + o_p(|\hat{\alpha}_n - \alpha_0|) + o_p(|\tilde{\alpha}_n - \alpha_0|), \end{aligned} \tag{2.5.20}$$

for $n \to \infty$. We will now look at $o_p(|\hat{\alpha}_n - \alpha_0|)$. Suppose $h \in o_p(|\hat{\alpha}_n - \alpha_0|)$ for $n \to \infty$. Then

$$\begin{aligned} h &= \frac{h}{|\hat{\alpha}_n - \alpha_0|}|\hat{\alpha}_n - \alpha_0| && (2.5.21) \\ &= |\hat{\alpha}_n - \alpha_0|o_p(1) && (2.5.22) \\ &\leq (|\hat{\alpha}_n - \tilde{\alpha}_n| + |\tilde{\alpha}_n - \alpha_0|)o_p(1) && (2.5.23) \\ &= |\hat{\alpha}_n - \tilde{\alpha}_n|o_p(1) + o_p(|\tilde{\alpha}_n - \alpha_0|). && (2.5.24) \end{aligned}$$

for $n \to \infty$. Therefore

$$\begin{aligned} 0 \leq {} & (\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)) + o_p(1))(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) \\ & + (\hat{\alpha}_n - \tilde{\alpha}_n)o_p(1) + o_p(|\tilde{\alpha}_n - \alpha_0|), \end{aligned} \tag{2.5.25}$$

for $n \to \infty$. By Theorem 2.9 we know that $\tilde{\alpha}_n - \alpha_0 = O_p\left(\frac{1}{n}\right)$ as $n \to \infty$. Thus it would make sense that

$$o_p(|\tilde{\alpha}_n - \alpha_0|) \subset o_p\left(\frac{1}{n}\right), \tag{2.5.26}$$

for $n \to \infty$. We prove this in the following lemma.

Lemma 2.16. *Suppose for the sequences of random variables A_n and B_n we have $A_n = o_p(B_n)$ and $B_n = O_p(1/n)$ respectively for $n \to \infty$. Then $A_n = o_p(1/n)$ for $n \to \infty$.*

Proof. Let $\epsilon_1, \epsilon_2 > 0$. Since $B_n = O_p(1/n)$ for $n \to \infty$, there exist $M > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}(n|B_n| \geq M) \leq \frac{\epsilon_2}{2}. \tag{2.5.27}$$

Since $A_n = o_p(B_n)$, there exists $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$ we have

$$\mathbb{P}\left(\left|\frac{A_n}{B_n}\right| \geq \frac{\epsilon_1}{M}\right) \leq \frac{\epsilon_2}{2}. \tag{2.5.28}$$

Define $N_0 := \max\{N_1, N_2\}$. If

$$|nB_n| < M \quad \text{and} \quad \left|\frac{A_n}{B_n}\right| < \epsilon_1/M. \tag{2.5.29}$$

then

$$n|A_n| < \epsilon_1. \tag{2.5.30}$$

Hence by using the Fréchet inequalities, we find that for $n \geq N_0$ we have

$$\mathbb{P}(n|A_n| < \epsilon_1) \geq \mathbb{P}\left(|nB_n| < M\right) + \mathbb{P}\left(\left|\frac{A_n}{B_n}\right| < \epsilon_1/M\right) - 1 \tag{2.5.31}$$

$$\geq 1 - \epsilon_2, \tag{2.5.32}$$

from which we conclude that indeed $A_n = o_p(1/n)$ for $n \to \infty$. ∎

By applying Theorem 2.9 and Lemma 2.16 we can see that Equation (2.5.26) does indeed hold. Thus we find that

$$\begin{aligned} 0 \leq\ & (\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)) + o_p(1))(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) \\ & + |\hat{\alpha}_n - \tilde{\alpha}_n| o_p(1) + o_p(1/n), \end{aligned} \tag{2.5.33}$$

for $n \to \infty$.

Now we want to investigate $\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)$. Using Theorem 2.1, which states that $\hat{\theta}_n$ is a consistent estimator for θ_0, and the continuous mapping theorem stated in Theorem B.3, it can clearly be seen that

$$\log(\beta(\alpha_0, \hat{\theta}_n)) - \log(\gamma(\alpha_0, \hat{\theta}_n)) = \log \frac{\beta_0}{\gamma_0} + o_p(1), \tag{2.5.34}$$

for $n \to \infty$. Therefore

$$0 \leq \left(\log \frac{\beta_0}{\gamma_0} + o_p(1)\right)(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) + (\hat{\alpha}_n - \tilde{\alpha}_n) o_p(1) + o_p(1/n), \tag{2.5.35}$$

for $n \to \infty$.

Since the sign of $\hat{\alpha}_n - \tilde{\alpha}_n$ is stochastically bounded by 1, we can clearly we can write

$$(\hat{\alpha}_n - \tilde{\alpha}_n)o_p(1) = |\hat{\alpha}_n - \tilde{\alpha}_n|o_p(1). \tag{2.5.36}$$

Hence

$$0 \leq \left(\log \frac{\beta_0}{\gamma_0} + o_p(1)\right)(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) + |\hat{\alpha}_n - \tilde{\alpha}_n|o_p(1) + o_p(1/n). \tag{2.5.37}$$

This expression is what will be used to prove Theorem 2.14 which states that $\hat{\alpha}_n - \tilde{\alpha}_n = o_p(1/n)$.

2.5.2 Proof of Theorem 2.14

We start by giving a heuristic overview of the proof of Theorem 2.14 and introducing a relevant lemma, which will be proven in the next subsection. Then we give the detailed proof.

The idea of the proof is that we bound $M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)$ both from above and below. Both of these bounds will include $|\tilde{\alpha}_n - \hat{\alpha}_n|$ and one of them includes $o_p(1/n)$. Then we use that the lower bound is bounded from above by the upper bound and do some more steps to get to the final conclusion. First we will use the following lemma, which will be proven in the next subsection.

Lemma 2.17. *For all $\epsilon > 0$ there exist $\delta, W > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have*

$$\mathbb{P}\left(\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} < W\right) > 1 - \epsilon, \tag{2.5.38}$$

and

$$\mathbb{P}(\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0) > 1 - \epsilon, \tag{2.5.39}$$

where

$$D_\delta^{(n)} := \{(\alpha, \theta) \in [0,1]^2 : \alpha \neq \tilde{\alpha}_n \wedge |\alpha - \alpha_0| \leq a_n \wedge |\theta - \theta_0| \leq \delta\}. \tag{2.5.40}$$

Lemma 2.17 basically says that

$$\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} < W, \tag{2.5.41}$$

and

$$\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0. \tag{2.5.42}$$

Theorem 2.1 states that $\hat{\theta}_n$ is a consistent estimator of θ_0. Hence $|\hat{\theta}_n - \theta_0| < \delta$ eventually. Since $\hat{\alpha}_n - \alpha_0 = o_p(a_n)$, we have $|\hat{\alpha}_n - \alpha_0| < a_n$ eventually. Thus $(\hat{\alpha}_n, \hat{\theta}_n) \in D_\delta^{(n)}$ eventually. But then

$$M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n) > 0, \tag{2.5.43}$$

and

$$\frac{|\tilde{\alpha}_n - \hat{\alpha}_n|}{M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)} < W. \tag{2.5.44}$$

However then also

$$\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| < M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n) \tag{2.5.45}$$

Equation (2.5.37) tells us that we can pick $A_n, B_n \in o_p(1)$ and $C_n \in o_p(1/n)$ such that

$$0 \leq \left(\log \frac{\beta_0}{\gamma_0} + A_n\right)(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) + |\hat{\alpha}_n - \tilde{\alpha}_n| B_n + C_n. \tag{2.5.46}$$

Since $\beta_0 > \gamma_0$ we have $\log \frac{\beta_0}{\gamma_0} + A_n > 0$ eventually. Thus

$$\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| \left(\log \frac{\beta_0}{\gamma_0} + A_n\right) < (M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)) \left(\log \frac{\beta_0}{\gamma_0} + A_n\right) \tag{2.5.47}$$

$$\leq |\hat{\alpha}_n - \tilde{\alpha}_n| B_n + C_n, \tag{2.5.48}$$

and hence

$$\frac{n}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| \left(\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)\right) < n \cdot C_n. \tag{2.5.49}$$

However we also have $\log \frac{\beta_0}{\gamma_0} + (A_n - B_n) > 0$ eventually. Therefore

$$n|\tilde{\alpha}_n - \hat{\alpha}_n| < \frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}. \tag{2.5.50}$$

We can show that the right hand side is $o_p(1)$, leading us to the conclusion that $\tilde{\alpha}_n - \hat{\alpha}_n = o_p(1/n)$.

Proof of Theorem 2.14. First we will bound $M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)$ from below by something positive. To do this, we will use Theorem 2.1, which states that $\hat{\theta}_n$ is consistent, and Lemma 2.17, which states that or all $\epsilon > 0$ there exist $\delta, W > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} < W\right) > 1 - \epsilon, \tag{2.5.51}$$

and

$$\mathbb{P}(\forall_{(\alpha,\theta) \in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0) > 1 - \epsilon, \tag{2.5.52}$$

where

$$D_\delta^{(n)} := \{(\alpha, \theta) \in [0,1]^2 : \alpha \neq \tilde{\alpha}_n \wedge |\alpha - \alpha_0| \leq a_n \wedge |\theta - \theta_0| \leq \delta\}. \tag{2.5.53}$$

Let $\epsilon_1, \epsilon_2 > 0$. By Lemma 2.17, we can choose $\delta, K > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}\big(\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\alpha,\theta) - M_n(\tilde{\alpha}_n,\theta) < 0\big) > 1 - \epsilon_1/24, \tag{2.5.54}$$

and

$$\mathbb{P}\left(\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n,\theta) - M_n(\alpha,\theta)} < K\right) > 1 - \epsilon_1/24. \tag{2.5.55}$$

Since $\hat{\theta}_n$ is a consistent estimator for θ_0, there exists $N_2 \in \mathbb{N}$ such that for $n \geq N_2$ we have

$$\mathbb{P}(|\hat{\theta}_n - \theta_0| < \delta) > 1 - \epsilon_1/24. \tag{2.5.56}$$

Since $\hat{\alpha}_n - \alpha_0 = o_p(a_n)$, there exists $N_3 \in \mathbb{N}$ such that for $n \geq N_3$ we have $\mathbb{P}(|\hat{\alpha}_n - \alpha_0| < a_n) > 1 - \epsilon/24$. If

$$|\hat{\alpha}_n - \alpha_0| < a_n, \quad \text{and} \quad |\hat{\theta}_n - \theta_0| < \delta, \tag{2.5.57}$$

then $(\hat{\alpha}_n, \hat{\theta}_n) \in D_\delta^{(n)}$. If additionally

$$\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\alpha,\theta) - M_n(\tilde{\alpha}_n,\theta) < 0, \tag{2.5.58}$$

then

$$M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n) > 0. \tag{2.5.59}$$

If also

$$\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n,\theta) - M_n(\alpha,\theta)} < K, \tag{2.5.60}$$

then

$$(\hat{\alpha}_n, \hat{\theta}_n) \in D_\delta^{(n)}, \quad \text{and} \quad \sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n,\theta) - M_n(\alpha,\theta)} < K, \tag{2.5.61}$$

and hence

$$\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| < M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n). \tag{2.5.62}$$

Thus we can use the Fréchet inequalities to find that for $n \geq \max\{N_1, N_2, N_3\}$

$$\mathbb{P}\left(\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| < M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)\right) > 1 - \frac{\epsilon_1}{6}. \tag{2.5.63}$$

We continue to bound $M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)$ from above. After finding the bounds we want to compare the bounds and create some bound on $n|\tilde{\alpha}_n - \hat{\alpha}_n|$. To find the upper bound for $M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)$, we consider Equation (2.5.37). It tells us that we can pick $A_n, B_n \in o_p(1)$ and $C_n \in o_p(1/n)$ such that

$$0 \leq \left(\log\frac{\beta_0}{\gamma_0} + A_n\right)(M_n(\hat{\alpha}_n, \hat{\theta}_n) - M_n(\tilde{\alpha}_n, \hat{\theta}_n)) + |\hat{\alpha}_n - \tilde{\alpha}_n|B_n + C_n. \tag{2.5.64}$$

Hence for all $n \in \mathbb{N}$ we have

$$\left(\log \frac{\beta_0}{\gamma_0} + A_n\right) (M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)) \leq |\hat{\alpha}_n - \tilde{\alpha}_n| B_n + C_n. \tag{2.5.65}$$

If

$$\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| < M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n) \tag{2.5.66}$$

and

$$\left(\log \frac{\beta_0}{\gamma_0} + A_n\right) > 0, \tag{2.5.67}$$

then

$$\frac{1}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| \left(\log \frac{\beta_0}{\gamma_0} + A_n\right) < (M_n(\tilde{\alpha}_n, \hat{\theta}_n) - M_n(\hat{\alpha}_n, \hat{\theta}_n)) \left(\log \frac{\beta_0}{\gamma_0} + A_n\right) \tag{2.5.68}$$

$$\leq |\hat{\alpha}_n - \tilde{\alpha}_n| B_n + C_n. \tag{2.5.69}$$

By taking $|\hat{\alpha}_n - \tilde{\alpha}_n| B_n$ to the left hand side and multiplying both sided by n, we also have

$$\frac{n}{K}|\tilde{\alpha}_n - \hat{\alpha}_n| \left(\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)\right) < n \cdot C_n. \tag{2.5.70}$$

If also

$$\log \frac{\beta_0}{\gamma_0} + (A_n - B_n) > 0, \tag{2.5.71}$$

then

$$n|\tilde{\alpha}_n - \hat{\alpha}_n| < \frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}. \tag{2.5.72}$$

We will now work out the probabilities.

Since by Assumption 2.2 we have $\beta_0 > \gamma_0$ and $A_n \in o_p(1)$ for $n \to \infty$, we can take $N_4 \in \mathbb{N}$ such that for $n \geq N_4$ we have that

$$\mathbb{P}\left(\left(\log \frac{\beta_0}{\gamma_0} + A_n\right) > 0\right) > 1 - \frac{\epsilon_1}{6}. \tag{2.5.73}$$

Since $(A_n - B_n) \in o_p(1)$ for $n \to \infty$ and we have assumed that $\beta_0 > \gamma_0$, there exists $N_5 \in \mathbb{N}$ such that for $n \geq N_5$ we have

$$\mathbb{P}\left(\log \frac{\beta_0}{\gamma_0} + (A_n - B_n) > 0\right) > 1 - \frac{\epsilon_1}{6}. \tag{2.5.74}$$

By the Fréchet inequalities we thus find that for $n \geq \max\{N_1, N_2, N_3, N_4, N_5\}$

$$\mathbb{P}\left(n|\tilde{\alpha}_n - \hat{\alpha}_n| < \frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}\right) > 1 - \frac{\epsilon_1}{2}. \tag{2.5.75}$$

All that we have left to do is to bound

$$\frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}, \tag{2.5.76}$$

by ϵ_2. To do this we shall first prove that $\frac{1}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)} = O_p(1)$ for $n \to \infty$.

Define $E_n := (A_n - B_n) \in o_p(1)$. Let $\epsilon' > 0$. Fix $\delta_1 \in \left(0, \log \frac{\beta_0}{\gamma_0}\right)$ and choose $N' \in \mathbb{N}$ such that for all $n \geq N'$ we have

$$\mathbb{P}(|E_n| \geq \delta_1) \leq \epsilon'. \tag{2.5.77}$$

Define

$$\delta_2 := \frac{1}{\log \frac{\beta_0}{\gamma_0} - \delta_1} > 0. \tag{2.5.78}$$

If

$$\left| \frac{1}{\log \frac{\beta_0}{\gamma_0} + E_n} \right| \geq \delta_2, \tag{2.5.79}$$

then by using the reverse triangle inequality we find that

$$\log \frac{\beta_0}{\gamma_0} - \delta_1 = \frac{1}{\delta_2} \tag{2.5.80}$$

$$\geq \left| \log \frac{\beta_0}{\gamma_0} + E_n \right| \tag{2.5.81}$$

$$\geq \left| \log \frac{\beta_0}{\gamma_0} - |E_n| \right| \tag{2.5.82}$$

$$\geq \log \frac{\beta_0}{\gamma_0} - |E_n|, \tag{2.5.83}$$

and therefore

$$\delta_1 \leq |E_n|. \tag{2.5.84}$$

Hence for $n \geq N'$ we have

$$\mathbb{P}\left(\left| \frac{1}{\log \frac{\beta_0}{\gamma_0} + E_n} \right| \geq \delta_2 \right) \leq \mathbb{P}(|E_n| \geq \delta_1) \leq \epsilon'. \tag{2.5.85}$$

Thus we can conclude that

$$\frac{1}{\log \frac{\beta_0}{\gamma_0} + E_n} = O_p(1), \tag{2.5.86}$$

for $n \to \infty$.

Since $C_n \in o_p(1/n)$ for $n \to \infty$, we can notice that $K \cdot n \cdot C_n \in o_p(1)$. Hence

$$\frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)} = o_p(1) O_p(1) = o_p(1). \tag{2.5.87}$$

Therefore we can choose $N_6 \in \mathbb{N}$ such that for $n \geq N_6$ we have

$$\mathbb{P}\left(\left|\frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}\right| \leq \epsilon_2\right) \geq 1 - \epsilon/2. \tag{2.5.88}$$

If

$$n|\tilde{\alpha}_n - \hat{\alpha}_n| < \frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}, \tag{2.5.89}$$

and

$$\left|\frac{K \cdot n \cdot C_n}{\log \frac{\beta_0}{\gamma_0} + (A_n - B_n)}\right| \leq \epsilon_2, \tag{2.5.90}$$

then

$$n|\tilde{\alpha}_n - \hat{\alpha}_n| < \epsilon_2. \tag{2.5.91}$$

Hence we can use the Fréchet inequalities to show that for $n \geq \max\{N_1, N_2, N_3, N_4, N_5, N_6\}$ we have

$$\mathbb{P}(n|\tilde{\alpha}_n - \hat{\alpha}_n| < \epsilon_2) \geq 1 - \epsilon, \tag{2.5.92}$$

from which we conclude that indeed

$$\tilde{\alpha}_n - \hat{\alpha}_n = o_p(1/n). \tag{2.5.93}$$

■

2.5.3 Proof of Lemma 2.17

In the proof of Theorem 2.14, we used Lemma 2.17, which states that for all $\epsilon > 0$ there exist $\delta, W > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} < W\right) > 1 - \epsilon, \tag{2.5.94}$$

and

$$\mathbb{P}(\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0) > 1 - \epsilon, \tag{2.5.95}$$

where

$$D_\delta^{(n)} := \{(\alpha, \theta) \in [0, 1]^2 : \alpha \neq \tilde{\alpha}_n \wedge |\alpha - \alpha_0| \leq a_n \wedge |\theta - \theta_0| \leq \delta\}. \tag{2.5.96}$$

Before talking about the proof of this lemma, we want to point out that Lemma 4 from the paper by Chernoff and Rubin[3], which corresponds to Lemma 2.14, is not correct. The lemma is roughly the same, but in the paper they state that δ does depend on ϵ.

To prove Lemma 2.14, we will use the following lemma, which will be proven in the next subsection.

Lemma 2.18. *We have*

$$\sup_{\substack{\alpha\neq\tilde{\alpha}_n \\ |\alpha-\alpha_0|\leq a_n}} \left(\frac{|\alpha-\tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n,\theta_0)-M_n(\alpha,\theta_0)} \right) = O_p(1), \tag{2.5.97}$$

for $n\to\infty$.

We first give a short introduction to the proof of Lemma 2.17. For any $\delta\geq 0$, we define

$$Y_n(\delta) := \sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n-\alpha|}{M_n(\tilde{\alpha}_n,\theta)-M_n(\alpha,\theta)}. \tag{2.5.98}$$

The proof relies on Lemma 2.18, which says that $Y_n(0)=O_p(1)$ as $n\to\infty$. The strategy of this proof is to first relate $Y_n(\delta)$ to $Y_n(0)$. We then bound $Y_n(0)$ by using Lemma 2.18. To continue we will find some $\delta>0$ such that we can bound $Y_n(\delta)$ by a multiple of $Y_n(0)$. Since $Y_n(0)$ was bounded already, we found a bound for $Y_n(\delta)$.

We now give the detailed proof of Lemma 2.17.

Proof of Lemma 2.17. Notice that $Y_n(\delta)$ is always positive.

Let $\epsilon>0$ be fixed. We first want to find a convenient expression for $M_n(\tilde{\alpha}_n,\theta)-M_n(\alpha,\theta)$, which is closer to what is used in $Y_n(0)$. Recall from Equation (2.3.1) that for any $\alpha,\theta\in(0,1)$ we have

$$M_n(\alpha,\theta) := F_n(\alpha)-F_n(\alpha_0)+\frac{(\gamma_0-\beta_0)(\alpha-\alpha_0)}{\log(\beta(\alpha_0,\theta))-\log(\gamma(\alpha_0,\theta))}. \tag{2.5.99}$$

Define $c:(0,1)\to\mathbb{R}$ by

$$c(\theta) := \frac{\beta_0-\gamma_0}{\log\beta(\alpha_0,\theta)-\log\gamma(\alpha_0,\theta)}. \tag{2.5.100}$$

Then

$$M_n(\alpha,\theta) = F_n(\alpha)-F_n(\alpha_0)-(\alpha-\alpha_0)c(\theta). \tag{2.5.101}$$

Therefore for any $\alpha,\theta\in(0,1)$ we have

$$\begin{aligned} M_n(\tilde{\alpha}_n,\theta)-M_n(\alpha,\theta) &= F_n(\tilde{\alpha}_n)-F_n(\alpha)-(\tilde{\alpha}_n-\alpha)c(\theta) & (2.5.102)\\ &= M_n(\tilde{\alpha}_n,\theta_0)-M_n(\alpha,\theta_0)+(\tilde{\alpha}_n-\alpha)(c(\theta_0)-c(\theta)). & (2.5.103)\end{aligned}$$

Now we bound $Y_n(0)$. We know from Lemma 2.18 that $Y_n(0)=O_p(1)$ for $n\to\infty$. Hence we can choose $N\in\mathbb{N}$ and $W>0$ such that for all $n\geq N$

$$\mathbb{P}\left(|Y_n(0)|<\frac{W}{2}\right) > 1-\epsilon. \tag{2.5.104}$$

Clearly $c(\theta)$ is continuous on its domain. Therefore we can choose $\delta>0$ such that

$$|\theta-\theta_0|<\delta \implies |c(\theta_0)-c(\theta)|<\frac{1}{W}. \tag{2.5.105}$$

Suppose that $Y_n(0) < W/2$. Then for any $\theta \in (\theta_0 - \delta, \theta_0 + \delta)$ we have

$$|c(\theta_0) - c(\theta)| < \frac{1}{W} < \frac{1}{2Y_n}. \tag{2.5.106}$$

Thus for all $(\alpha, \theta) \in D_\delta^{(n)}$

$$\begin{aligned} M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) &= M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0) + (\tilde{\alpha}_n - \alpha)(c(\theta_0) - c(\theta)) && (2.5.107) \\ &\geq M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0) - |\tilde{\alpha}_n - \alpha| \cdot |c(\theta_0) - c(\theta)| && (2.5.108) \\ &\geq M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0) - \frac{|\tilde{\alpha}_n - \alpha|}{2Y_n(0)}. && (2.5.109) \end{aligned}$$

However by definition of $Y_n(0)$, we know that for any $(\alpha, \theta) \in D_\delta^{(n)}$ we also have

$$\frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \leq Y_n(0). \tag{2.5.110}$$

By definition of $\tilde{\alpha}_n$, we know that for any $(\alpha, \theta) \in D_\delta^{(n)}$

$$M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0) > 0. \tag{2.5.111}$$

By recalling that $Y_n(0) > 0$ and using Equations (2.5.110) and (2.5.111), it can be seen that for any $(\alpha, \theta) \in D_\delta^{(n)}$

$$M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0) \geq \frac{|\tilde{\alpha}_n - \alpha|}{Y_n(0)}. \tag{2.5.112}$$

Plugging this back into Equation (2.5.109) yields that for all $(\alpha, \theta) \in D_\delta^{(n)}$

$$M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) \geq \frac{|\tilde{\alpha}_n - \alpha|}{2Y_n(0)} > 0. \tag{2.5.113}$$

But then for all $(\alpha, \theta) \in D_\delta^{(n)}$ we find that

$$\frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} \leq 2Y_n(0), \tag{2.5.114}$$

and thus

$$Y_n(\delta) \leq 2Y_n(0) \leq W. \tag{2.5.115}$$

Hence for $n \geq N$ we have

$$\mathbb{P}(|Y_n(\delta)| < W) \geq \mathbb{P}(Y_n(0) < W/2) > 1 - \epsilon. \tag{2.5.116}$$

Since $Y_n(0) < W/2$ also implies Equation (2.5.113), we also have that for $n \geq N$

$$\mathbb{P}\left(\forall_{(\alpha,\theta) \in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0\right) > 1 - \epsilon. \tag{2.5.117}$$

Thus we conclude that indeed for all $\epsilon > 0$ there exist $\delta, W > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{D_\delta^{(n)}} \frac{|\tilde{\alpha}_n - \alpha|}{M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta)} < W\right) > 1 - \epsilon, \tag{2.5.118}$$

and

$$\mathbb{P}(\forall_{(\alpha,\theta)\in D_\delta^{(n)}} : M_n(\tilde{\alpha}_n, \theta) - M_n(\alpha, \theta) > 0) > 1 - \epsilon. \tag{2.5.119}$$

■

2.5.4 Proof of Lemma 2.18

We still need to show Lemma 2.18, which was used in the proof of Lemma 2.17. This lemma states that

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| \leq a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)}\right) = O_p(1), \tag{2.5.120}$$

for $n \to \infty$.

We start by giving a short overview of the proof. Then we will give the proof in full detail.

Recall that in the proof of Theorem 2.9 we used a lemma that was stronger than what we needed in that proof. This lemma was Lemma 2.10. It was mentioned that we proved this stronger version because it would be useful later on. It will be crucial to the proof of this lemma. This lemma says that for all $\epsilon > 0$ there exist $K, k > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k\right) > 1 - \epsilon. \tag{2.5.121}$$

Now we can split the supremum that we want to bound in

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| \leq a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)}\right), \tag{2.5.122}$$

into two parts. These parts are for $|\alpha - \alpha_0| < K/n$ and $K/n < |\alpha - \alpha_0| < a_n$. We start by bounding the supremum over the latter. Because of Lemma 2.10 we have $|\tilde{\alpha}_n - \alpha_0| < K/n$. A few simple steps will lead us to the bound on this supremum. The more involved part of the proof is bounding the supremum over $|\alpha - \alpha_0| < K/n$. To do this we must first introduce some new notation for the observations. Whenever we talk about a sample of size n, we order the observations and give them the labels

$$\ldots, x_{-2}^{(n)}, x_{-1}^{(n)}, x_1^{(n)}, x_2^{(n)}, \ldots \tag{2.5.123}$$

such that

$$\cdots < x_{-2}^{(n)} < x_{-1}^{(n)} < x_0^{(n)} < x_1^{(n)} < x_2^{(n)} < \cdots , \tag{2.5.124}$$

where

$$x_0^{(n)} := \alpha_0. \tag{2.5.125}$$

We call these the order statistics of the observations.

We will first show that there is some $U \in \mathbb{N}$ such that with great probability for n large enough we have $x_{-U}^{(n)} < \alpha_0 - K/n$ and $x_U^{(n)} > \alpha_0 + K/n$. What this means is that there will be less than U observations on the left and right hand side of α_0 which lie in the interval $[\alpha_0 - K/n, \alpha_0 + K/n]$. Then we will show that $\tilde{\alpha}_n$ is one of the observations that do lie in this interval. We use all of these things to show that the supremum we are considering is bounded by the sequence of random variables

$$Z_n := \sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right| . \tag{2.5.126}$$

Then we use the following lemma, which will be proven in the next subsection.

Lemma 2.19. *For any fixed $U \in \mathbb{N}$ the sequence of random variables*

$$\sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right| , \tag{2.5.127}$$

converges in distribution.

Since convergent sequences of random variables are stochastically bounded, Lemma 2.19 tells us that Z_n is stochastically bounded too. But then the supremum that is bounded by Z_n is also bounded, leading us to the final conclusion. Now we give the detailed proof.

Proof of Lemma 2.18. Define

$$Y_n := \sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| \leq a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) . \tag{2.5.128}$$

Fix $\epsilon > 0$. By using Lemma 2.10 there exist $K, k > 0$ and $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k \right) > 1 - \epsilon/3. \tag{2.5.129}$$

This will be useful in analysing Y_n, because it allows us to split the supremum over $\{\alpha \in (0,1) : \alpha \neq \tilde{\alpha}_n \wedge |\alpha - \alpha_0| \leq a_n\}$ into two parts. First we will bound

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) . \tag{2.5.130}$$

Afterwards we continue with

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| < K/n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right). \tag{2.5.131}$$

In order to bound the former, we will use Equation (2.5.129). Suppose

$$\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k. \tag{2.5.132}$$

By definition of the QMLE and M_n we have

$$M_n(\tilde{\alpha}_n, \theta_0) \geq M_n(\alpha_0, \theta_0) = 0. \tag{2.5.133}$$

Therefore

$$|\tilde{\alpha}_n - \alpha_0| < \frac{K}{n}. \tag{2.5.134}$$

If $\alpha \in (0, 1)$ such that $K/n < |\alpha - \alpha_0| < a_n$, then

$$\frac{-M_n(\alpha, \theta_0)}{k} > |\alpha - \alpha_0|. \tag{2.5.135}$$

Since $M_n(\tilde{\alpha}_n, \theta_0) \geq 0$, we find that

$$\frac{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)}{k} > |\alpha - \alpha_0|. \tag{2.5.136}$$

As this holds for all $\alpha \in (0, 1)$ such that $K/n < |\alpha - \alpha_0| < a_n$, it follows by Equation (2.5.134) that under the assumption of Equation (2.5.132)

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) \leq \sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{k|\alpha - \alpha_0|} \right) \tag{2.5.137}$$

$$\leq \sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \left(\frac{|\alpha - \alpha_0| + |\alpha_0 - \tilde{\alpha}_n|}{k|\alpha - \alpha_0|} \right) \tag{2.5.138}$$

$$= \frac{1}{k} + \sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \frac{|\alpha_0 - \tilde{\alpha}_n|}{k|\alpha - \alpha_0|} \tag{2.5.139}$$

$$\leq \frac{1}{k} + \frac{\frac{K}{n}}{\frac{K \cdot k}{n}} \tag{2.5.140}$$

$$= \frac{2}{k}. \tag{2.5.141}$$

Hence for $n \geq N_1$ we have

$$\mathbb{P}\left(\sup_{\substack{\alpha\neq\tilde{\alpha}_n \\ K/n<|\alpha-\alpha_0|<a_n}}\left(\frac{|\alpha-\tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n,\theta_0)-M_n(\alpha,\theta_0)}\right)\leq\frac{2}{k}\right)\geq\mathbb{P}\left(\sup_{K/n<|\alpha-\alpha_0|<a_n}\frac{M_n(\alpha,\theta_0)}{|\alpha-\alpha_0|}<-k\right) \tag{2.5.142}$$

$$> 1-\epsilon/3. \tag{2.5.143}$$

We will now consider

$$\sup_{\substack{\alpha\neq\tilde{\alpha}_n \\ |\alpha-\alpha_0|<K/n}}\left(\frac{|\alpha-\tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n,\theta_0)-M_n(\alpha,\theta_0)}\right). \tag{2.5.144}$$

Bounding this will be more technical. We will use the steps as mentioned in the outline of the proof.

We start by showing that there is some $U \in \mathbb{N}$ such that with great probability for n large enough we have $x_{-U}^{(n)} < \alpha_0 - K/n$ and $x_U^{(n)} > \alpha_0 + K/n$. By the Fréchet inequalities, we know that for any $U \in \mathbb{N}$ we have

$$\mathbb{P}\left(x_{-U}^{(n)}<\alpha_0-\frac{K}{n}, x_U^{(n)}>\alpha_0+\frac{K}{n}\right)\geq\mathbb{P}\left(x_{-U}^{(n)}<\alpha_0-\frac{K}{n}\right)+\mathbb{P}\left(x_U^{(n)}>\alpha_0+\frac{K}{n}\right)-1. \tag{2.5.145}$$

However since that is the same as the probability of there being less than U observations in each $(\alpha_0, \alpha_0 + K/n)$ and $(\alpha_0 - K/n, \alpha_0)$, we find that

$$\mathbb{P}\left(x_{-U}^{(n)}<\alpha_0-\frac{K}{n}\right)=\mathbb{P}\left(n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)<U\right), \tag{2.5.146}$$

and

$$\mathbb{P}\left(x_U^{(n)}>\alpha_0+\frac{K}{n}\right)=\mathbb{P}\left(n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)<U\right). \tag{2.5.147}$$

Notice that since they count the amount of times observations lie within the intervals, we have

$$n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)\sim\text{Binom}\left(n,\frac{K\beta_0}{n}\right), \tag{2.5.148}$$

and

$$n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)\sim\text{Binom}\left(n,\frac{K\gamma_0}{n}\right). \tag{2.5.149}$$

It is well known that for any fixed $q \in \mathbb{R}^+$

$$\text{Binom}\left(n,\frac{q}{n}\right)\xrightarrow{p}\text{Pois}(q). \tag{2.5.150}$$

Thus

$$n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)\xrightarrow{d}\text{Pois}(K\beta_0), \tag{2.5.151}$$

and

$$n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)\xrightarrow{d}\text{Pois}(K\gamma_0). \tag{2.5.152}$$

Let $U_1, U_2 \in \mathbb{N}$ such that

$$\mathbb{P}(\text{Pois}(K\beta_0) < U_1) > 1-\frac{\epsilon}{12}, \tag{2.5.153}$$

and

$$\mathbb{P}(\text{Pois}(K\gamma_0) < U_2) > 1-\frac{\epsilon}{12}. \tag{2.5.154}$$

Fix $U := \max\{U_1, U_2\}$. Since

$$\begin{aligned}\mathbb{P}\left(n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)<U\right) &\geq \mathbb{P}\left(n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)<U_1\right) && (2.5.155)\\ &\to \mathbb{P}(\text{Pois}(K\beta_0)<U_1) && (2.5.156)\\ &> 1-\frac{\epsilon}{12}, && (2.5.157)\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)<U\right) &\geq \mathbb{P}\left(n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)<U_2\right) && (2.5.158)\\ &\to \mathbb{P}(\text{Pois}(K\gamma_0)<U_2) && (2.5.159)\\ &> 1-\frac{\epsilon}{12}, && (2.5.160)\end{aligned}$$

we can choose $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$ we have

$$\mathbb{P}\left(n\left(F_n(\alpha_0)-F_n\left(\alpha_0-\frac{K}{n}\right)\right)<U\right)>1-\frac{\epsilon}{6}, \tag{2.5.161}$$

and

$$\mathbb{P}\left(n\left(F_n\left(\alpha_0+\frac{K}{n}\right)-F_n(\alpha_0)\right)<U\right)>1-\frac{\epsilon}{6}. \tag{2.5.162}$$

Then by Equation (2.5.145), we find that for all $n \geq N_2$

$$\mathbb{P}\left(x_{-U}^{(n)}<\alpha_0-\frac{K}{n}, x_U^{(n)}>\alpha_0+\frac{K}{n}\right)>1-\frac{\epsilon}{3}. \tag{2.5.163}$$

Now we have a bound on the amount of observations on the left and right hand side of α_0 which lie in the interval $[\alpha_0 - K/n, \alpha_0 + K/n]$. Now we will show that $\tilde{\alpha}_n$ is one of the observations that do lie in this interval. Recall that

$$\tilde{\alpha}_n = \operatorname*{arg\,max}_{\alpha\in[\alpha_0-a_n,\alpha_0+a_n]} M_n(\alpha,\theta_0). \tag{2.5.164}$$

Since

$$M_n(\alpha, \theta_0) = F_n(\alpha) - F_n(\alpha_0) + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0}(\alpha - \alpha_0), \tag{2.5.165}$$

and

$$\frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} < 0, \tag{2.5.166}$$

we can see that $M_n(\alpha, \theta_0)$ decreases between observations, but jumps up at the observations. Hence $\tilde{\alpha}_n$ is equal to one of the observations in $[\alpha_0 - a_n, \alpha_0 + a_n]$ or one of the boundary points of this interval, $\alpha_0 \pm a_n$. If again we suppose that

$$\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k, \tag{2.5.167}$$

then, as mentioned before in Equation (2.5.134),

$$|\tilde{\alpha}_n - \alpha_0| < \frac{K}{n} < a_n. \tag{2.5.168}$$

Hence if Equation (2.5.167) holds, then $\tilde{\alpha}_n$ is equal to an observation with a distance to α_0 which is less than K/n.

We will now move on to finding a bounding sequence of random variables. In addition to Equation (2.5.167) we also suppose that

$$x_{-U}^{(n)} < \alpha_0 - \frac{K}{n}, \quad \text{and} \quad x_U^{(n)} > \alpha_0 + \frac{K}{n}. \tag{2.5.169}$$

For all $n \in \mathbb{N}$ define

$$Z_n := \sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right|. \tag{2.5.170}$$

We will now show that for all $n \in \mathbb{N}$

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| < K/n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) \leq Z_n, \tag{2.5.171}$$

under assumption of Equations (2.5.167) and (2.5.169).

Fix $n \in \mathbb{N}$. Since $\tilde{\alpha}_n$ is one of the observations in the interval $(\alpha_0 - K/n, \alpha_0 + K/n)$, there is $j_0 \in \mathbb{N} \cap [-U, U]$ such that $\tilde{\alpha}_n = x_{j_0}^{(n)}$. Since Z_n is the supremum over all unequal combinations of j_1 and j_2, we find that

$$Z_n \geq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \left| \frac{x_j^{(n)} - x_{j_0}^{(n)}}{M_n(x_j^{(n)}, \theta_0) - M_n(x_{j_0}^{(n)}, \theta_0)} \right| = \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left| \tilde{\alpha}_n - x_j^{(n)} \right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)}. \tag{2.5.172}$$

Let $\alpha \in (\alpha_0 - K/n, \alpha_0 + K/n) \setminus \{\tilde{\alpha}_n\}$. If α is one of the observations or α_0, then obviously

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \leq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left|\tilde{\alpha}_n - x_j^{(n)}\right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)}. \tag{2.5.173}$$

We will thus consider α that are not one of the observations or α_0. Let $j_1 \in \mathbb{N} \cap [-U, U-1]$ such that $\alpha \in (x_{j_1}^{(n)}, x_{j_1+1}^{(n)})$. We will consider three separate cases. The first case is the case where $j_1 < j_0$. The second case is $j_1 > j_0$. The last case is $j_1 = j_0$. In all cases we have

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} = \frac{|\alpha - \tilde{\alpha}_n|}{F_n(\tilde{\alpha}_n) - F_n(\alpha) + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0}(\tilde{\alpha}_n - \alpha)} \tag{2.5.174}$$

$$= \frac{1}{\frac{F_n(\tilde{\alpha}_n) - F_n(\alpha)}{|\alpha - \tilde{\alpha}_n|} + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \cdot \operatorname{sgn}(\tilde{\alpha}_n - \alpha)}. \tag{2.5.175}$$

We will now prove that in each case Equation (2.5.173) holds.

1. In the first case we assume that $j_1 < j_0$. Then $\alpha < \tilde{\alpha}_n$. For $\alpha < \tilde{\alpha}_n$ we can see that

$$\frac{F_n(\tilde{\alpha}_n) - F_n(\alpha)}{|\alpha - \tilde{\alpha}_n|} + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \cdot \operatorname{sgn}(\tilde{\alpha}_n - \alpha) \tag{2.5.176}$$

increases between observations and decreases with a jump at the observations themselves. Therefore

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \tag{2.5.177}$$

decreases between observations and increases with a jump at the observations themselves. Hence

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \leq \frac{|x_{j_1}^{(n)} - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \tag{2.5.178}$$

$$\leq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left|\tilde{\alpha}_n - x_j^{(n)}\right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)}. \tag{2.5.179}$$

2. In the second case we assume that $j_1 > j_0$. Then $\alpha > x_{j_0+1}^{(n)}$. For $\alpha > x_{j_0+1}^{(n)}$ we can see that

$$F_n(\tilde{\alpha}_n) - F_n(\alpha) < 0 \tag{2.5.180}$$

and therefore Equation (2.5.176) increases between observations and decreases with a jump at the observations themselves. Thus Equation (2.5.177) decreases between observations and increases with a jump at the observations themselves. Hence

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \leq \frac{|x_{j_1}^{(n)} - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \tag{2.5.181}$$

$$\leq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left|\tilde{\alpha}_n - x_j^{(n)}\right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)}. \tag{2.5.182}$$

3. In the last case we assume that $j_1 = j_0$. Then $x_{j_0}^{(n)} < \alpha < x_{j_0+1}^{(n)}$ and hence

$$\frac{F_n(\tilde{\alpha}_n) - F_n(\alpha)}{|\alpha - \tilde{\alpha}_n|} + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \cdot \operatorname{sgn}(\tilde{\alpha}_n - \alpha) = \frac{\beta_0 - \gamma_0}{\log \beta_0 - \log \gamma_0}. \tag{2.5.183}$$

The difference with the previous two cases is that we can not consider the observationn to the left of α, as this is the QMLE, which is not in the supremum. However, this time there is no decrease between observations, but it is constant. The jump at the next observation therefore always gives us an upper bound. This can be written more precisely as

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} = \frac{\log \beta_0 - \log \gamma_0}{\beta_0 - \gamma_0} \tag{2.5.184}$$

$$\leq \frac{1}{\frac{\beta_0 - \gamma_0}{\log \beta_0 - \log \gamma_0} - \frac{1}{n|\tilde{\alpha}_n - x_{j_0+1}^{(n)}|}} \tag{2.5.185}$$

$$= \frac{|x_{j_1+1}^{(n)} - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_{j_1+1}^{(n)}, \theta_0)} \tag{2.5.186}$$

$$\leq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left|\tilde{\alpha}_n - x_j^{(n)}\right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)}. \tag{2.5.187}$$

We can conclude that for all $\alpha \in (\alpha_0 - K/n, \alpha_0 + K/n) \setminus \{\alpha_0\}$ we have

$$\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \leq \sup_{\substack{-U \leq j \leq U \\ j \neq j_0}} \frac{\left|\tilde{\alpha}_n - x_j^{(n)}\right|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(x_j^{(n)}, \theta_0)} \leq Z_n, \tag{2.5.188}$$

and therefore indeed for all $n \in \mathbb{N}$

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| < K/n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) \leq Z_n. \tag{2.5.189}$$

To complete the proof of the lemma, we want to find $N \in \mathbb{N}$ and $\Xi \in \mathbb{R}$ such that for $n \geq N$ we have

$$\mathbb{P}(Z_n < \Xi) > 1 - \frac{\epsilon}{3}. \tag{2.5.190}$$

By Lemma 2.19, we know that Z_n converges in probability to some random variable Z. We can choose $\Xi \in \mathbb{R}$ such that

$$\mathbb{P}(Z < \Xi) > 1 - \frac{\epsilon}{6}. \tag{2.5.191}$$

But then we can choose $N_3 \in \mathbb{N}$ such that for all $n \geq N_3$ we have

$$\mathbb{P}(Z_n < \Xi) > 1 - \frac{\epsilon}{3}. \tag{2.5.192}$$

If

$$\sup_{K/n < |\alpha - \alpha_0| < a_n} \frac{M_n(\alpha, \theta_0)}{|\alpha - \alpha_0|} < -k \quad \text{and} \quad Z_n < \Xi, \tag{2.5.193}$$

and

$$x_{-U}^{(n)} < \alpha_0 - \frac{K}{n}, \quad \text{and} \quad x_U^{(n)} > \alpha_0 + \frac{K}{n}, \tag{2.5.194}$$

then

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ K/n < |\alpha - \alpha_0| < a_n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) \leq \frac{2}{k}, \tag{2.5.195}$$

and

$$\sup_{\substack{\alpha \neq \tilde{\alpha}_n \\ |\alpha - \alpha_0| < K/n}} \left(\frac{|\alpha - \tilde{\alpha}_n|}{M_n(\tilde{\alpha}_n, \theta_0) - M_n(\alpha, \theta_0)} \right) \leq Z_n < \Xi. \tag{2.5.196}$$

Therefore we also have

$$Y_n \leq \max \left\{ \frac{2}{k}, \Xi \right\}. \tag{2.5.197}$$

By using the Fréchet inequalities, we thus find that for $n \geq \max\{N_1, N_2, N_3\}$

$$\mathbb{P}\left(Y_n \leq \max \left\{ \frac{2}{k}, \Xi \right\} \right) \geq 1 - \epsilon, \tag{2.5.198}$$

from which we conclude that $Y_n = O_p(1)$ as $n \to \infty$. ■

2.5.5 Proof of Lemma 2.19

The last lemma that has to be proven in this section is Lemma 2.19, which states that for any fixed $U \in \mathbb{N}$ the sequence of random variables

$$\sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right|, \tag{2.5.199}$$

converges in distribution.

We first give a short overview of the proof. Then we provide the detailed proof.

To start we define

$$Z_n := \sup_{-U \le j_1 < j_2 \le U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right|. \tag{2.5.200}$$

We want to rewrite Z_n. To do this we define

$$y_i^{(n)} := n(x_i^{(n)} - x_{i-1}^{(n)}), \tag{2.5.201}$$

and

$$j^* := \begin{cases} j, & \text{if } j \ge 0, \\ j+1, & \text{if } j < 0. \end{cases} \tag{2.5.202}$$

After a few computational steps we find that

$$Z_n = \sup_{-U \le j_1 < j_2 \le U} \left| \frac{\sum_{i=j_1+1}^{j_2} y_i^{(n)}}{j_2^* - j_1^* + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \sum_{i=j_1+1}^{j_2} y_i^{(n)}} \right|. \tag{2.5.203}$$

We then define $\Psi : \mathbb{R}^{2U} \to \mathbb{R}$ by

$$\Psi(z_{-U+1}, \dots z_{-1}, z_0, z_1, \dots, z_U) = \sup_{-U \le j_1 < j_2 \le U} \left| \frac{\sum_{i=j_1+1}^{j_2} z_i}{j_2^* - j_1^* + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \sum_{i=j_1+1}^{j_2} z_i} \right|. \tag{2.5.204}$$

Then $Z_n = \Psi(y_{-U+1}^{(n)}, \dots, y_U^{(n)})$. We show that the joint distribution $(y_{-U+1}^{(n)}, \dots, y_U^{(n)})$ converges in distribution and find the limiting distribution. Then we will show that Ψ is continuous. Finally we will use the continuous mapping theorem to show that Z_n converges in distribution.

Proof of Lemma 2.19. We first rewrite Z_n. For all i define

$$y_i^{(n)} := n(x_i^{(n)} - x_{i-1}^{(n)}). \tag{2.5.205}$$

Then

$$x_{j_2}^{(n)} - x_{j_1}^{(n)} = \frac{1}{n} \sum_{i=j_1+1}^{j_2} y_i^{(n)}. \tag{2.5.206}$$

Therefore

$$Z_n = \sup_{-U \le j_1 < j_2 \le U} \left| \frac{x_{j_2}^{(n)} - x_{j_1}^{(n)}}{M_n(x_{j_2}^{(n)}, \theta_0) - M_n(x_{j_1}^{(n)}, \theta_0)} \right| \tag{2.5.207}$$

$$= \sup_{-U \le j_1 < j_2 \le U} \left| \frac{\frac{1}{n} \sum_{i=j_1+1}^{j_2} y_i^{(n)}}{F_n(x_{j_2}^{(n)}) - F_n(x_{j_1}^{(n)}) + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} (x_{j_2}^{(n)} - x_{j_1}^{(n)})} \right| \tag{2.5.208}$$

$$= \sup_{-U \le j_1 < j_2 \le U} \left| \frac{\sum_{i=j_1+1}^{j_2} y_i^{(n)}}{n(F_n(x_{j_2}^{(n)}) - F_n(x_{j_1}^{(n)})) + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \sum_{i=j_1+1}^{j_2} y_i^{(n)}} \right|. \tag{2.5.209}$$

Also define

$$j^* := \begin{cases} j, & \text{if } j \geq 0, \\ j+1, & \text{if } j < 0. \end{cases} \tag{2.5.210}$$

Since $n(F_n(x_{j_2}^{(n)}) - F_n(x_{j_1}^{(n)}))$ is equal to the amount of observations in the interval $(x_{j_1}^{(n)}, x_{j_2}^{(n)}]$ and α_0 is not an observation,

$$n(F_n(x_{j_2}^{(n)}) - F_n(x_{j_1}^{(n)})) = \begin{cases} j_2 - j_1 - 1, & \text{if } j_1 < 0 \leq j_2, \\ j_2 - j_1, & \text{otherwise} \end{cases} \tag{2.5.211}$$

$$= j_2^* - j_1^*. \tag{2.5.212}$$

Therefore we can write

$$Z_n = \sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{\sum_{i=j_1+1}^{j_2} y_i^{(n)}}{j_2^* - j_1^* + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \sum_{i=j_1+1}^{j_2} y_i^{(n)}} \right|. \tag{2.5.213}$$

Thus when we define $\Psi : \mathbb{R}^{2U} \to \mathbb{R}$ by

$$\Psi(z_{-U+1}, \ldots z_{-1}, z_0, z_1, \ldots, z_U) = \sup_{-U \leq j_1 < j_2 \leq U} \left| \frac{\sum_{i=j_1+1}^{j_2} z_i}{j_2^* - j_1^* + \frac{\gamma_0 - \beta_0}{\log \beta_0 - \log \gamma_0} \sum_{i=j_1+1}^{j_2} z_i} \right|, \tag{2.5.214}$$

we have

$$Z_n = \Psi(y_{-U+1}^{(n)}, \ldots, y_U^{(n)}). \tag{2.5.215}$$

We are now going show that the joint distribution $(y_{-U+1}^{(n)}, \ldots, y_U^{(n)})$ converges in distribution and find the limiting distribution. Then we will show that Ψ is continuous. Finally we will use the continuous mapping theorem to show that Z_n converges in distribution.

To show that the joint distribution $(y_{-U+1}^{(n)}, \ldots, y_U^{(n)})$ converges in distribution and find the limiting distribution, we want to use the available knowledge about order statistics of uniform random variables. To that end we first define the random variables

$$Q_i = \begin{cases} \beta_0 X_i + 1 - \theta_0, & \text{if } X_i < \alpha_0, \\ \gamma_0 (X_i - \alpha_0), & \text{otherwise.} \end{cases} \tag{2.5.216}$$

This makes sense as Q_i can be seen as flipping the left and right hand side of α_0 and stretching them to create a uniform distribution. This will be useful to the proof, since the first and last U observations are the observations we are interested in and the order statistics of uniform distributions have convenient properties that we can exploit.

Notice that $Q_i \in (0, 1-\theta_0)$ if and only if $X_i \in (\alpha_0, 1)$. Similarly notice that $Q_i \in (1-\theta_0, 1)$ if and only if $X_i \in (0, \alpha_0)$ Thus for $x \in [0, 1-\theta_0]$

$$\mathbb{P}(Q_i \leq x) = \mathbb{P}\left(X_i \in \left[\alpha_0, \alpha_0 + \frac{x}{\gamma_0}\right]\right) = \left(\alpha_0 + \frac{x}{\gamma_0} - \alpha_0\right)\gamma_0 = x, \tag{2.5.217}$$

and for $x \in [1-\theta_0, 1]$ we have

$$\mathbb{P}(Q_i \leq x) = \mathbb{P}\left(X_i \in [\alpha_0, 1] \cup \left[0, \frac{x-1+\theta_0}{\beta_0}\right]\right) \tag{2.5.218}$$

$$= (1-\alpha_0)\gamma_0 + \frac{x-1+\theta_0}{\beta_0}\beta_0 \tag{2.5.219}$$

$$= x. \tag{2.5.220}$$

Therefore

$$Q_i \sim \mathrm{Unif}([0,1]). \tag{2.5.221}$$

Whenever we have a sample of size n, we order the observations to the get the order statistics for Q_i. We denote these as

$$q_1^{(n)}, \dots, q_n^{(n)}, \tag{2.5.222}$$

such that

$$0 =: q_0^{(n)} < q_1^{(n)} < \cdots < q_n^{(n)} < q_{n+1}^{(n)} := 1. \tag{2.5.223}$$

For $i \in \{1, \dots, n+1\}$ and $n \in \mathbb{N}$ define

$$v_i^{(n)} := n(q_i^{(n)} - q_{i-1}^{(n)}). \tag{2.5.224}$$

Then with probability 1 there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$q_i^{(n)} = \gamma_0(x_i^{(n)} - \alpha_0), \quad \text{for } i \in \{0, \dots, U\}, \tag{2.5.225}$$

and

$$q_{n+1+i}^{(n)} = \beta_0 x_i^{(n)} + 1 - \theta_0, \quad \text{for } i \in \{-U+1, \dots, 0\}. \tag{2.5.226}$$

Therefore when $n \geq N$ we have

$$v_i^{(n)} = \gamma_0 y_i^{(n)} \quad \text{for } i \in \{1, \dots, U\}, \tag{2.5.227}$$

and

$$v_{n+1+i}^{(n)} = \beta_0 y_i^{(n)} \quad \text{for } i \in \{-U+1, \dots, 0\}. \tag{2.5.228}$$

It is well known that

$$(v_1^{(n)}, \dots, v_{n+1}^{(n)}) \sim \left(\frac{V_1}{\overline{V}_{n+1}}, \dots, \frac{V_{n+1}}{\overline{V}_{n+1}}\right), \tag{2.5.229}$$

where all V_i are i.i.d. exponential distributions with mean 1 and $\overline{V}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} V_i$. This statement is formulated well in Equation 1 of [6] and proven in Section 4.1 of [7]. We consider

$$(v_1^{(n)}, \dots, v_U^{(n)}, v_{n-U+2}^{(n)}, \dots, v_{n+1}^{(n)}) \sim \left(\frac{V_1}{\overline{V}_{n+1}}, \dots, \frac{V_{2U}}{\overline{V}_{n+1}}\right), \tag{2.5.230}$$

where all V_i for $i \in \{1,\dots,n+1\}$ are i.i.d. exponential distributions with mean 1 and $\overline{V}_{n+1} = \frac{1}{n}\sum_{i=1}^{n+1} V_i$. By the law of large numbers we find that $\overline{V}_{n+1} \xrightarrow{a.s.} 1$. Thus by the continuous mapping theorem

$$(v_1^{(n)},\dots,v_U^{(n)},v_{n-U+2}^{(n)},\dots,v_{n+1}^{(n)}) \xrightarrow{d} (V_1,\dots,V_{2U}). \tag{2.5.231}$$

But then

$$(y_{-U+1}^{(n)},\dots,y_U^{(n)}) = \left(\frac{v_{n-U+2}^{(n)}}{\beta_0}\dots,\frac{v_{n+1}^{(n)}}{\beta_0},\frac{v_1^{(n)}}{\gamma_0}\dots,\frac{v_U^{(n)}}{\gamma_0}\right) \tag{2.5.232}$$

$$\xrightarrow{d} \left(\frac{V_{U+1}}{\beta_0}\dots,\frac{V_{2U}}{\beta_0},\frac{V_1}{\gamma_0}\dots,\frac{V_U}{\gamma_0}\right). \tag{2.5.233}$$

So $y_{-U+1}^{(n)},\dots,y_0^{(n)}$ converge in distribution to i.i.d. exponential distributions with mean $1/\beta_0$ and $y_1^{(n)},\dots,y_U^{(n)}$ converge in distribution to i.i.d. exponential distributions with mean $1/\gamma_0$. The joint limiting distribution is mutually independent. Let $(y_{-U+1},\dots,y_U)$ be the limiting distribution.

We will continue by showing that Ψ is continuous. We can consider Ψ with domain $\mathbb{R}^{2U}$. Define $Q := \{j_1, j_2 \in [-U,U] \cap \mathbb{N} : j_1 < j_2\}$ Let $\Theta : \mathbb{R}^{2U} \times Q \to \mathbb{R}$ be defined by

$$\Theta(z_{-U+1},\dots,z_U,j_1,j_2) := \left|\frac{\sum_{i=j_1+1}^{j_2} z_i}{j_2^* - j_1^* + \frac{\gamma_0-\beta_0}{\log\beta_0-\log\gamma_0}\sum_{i=j_1+1}^{j_2} z_i}\right|. \tag{2.5.234}$$

Then

$$\Psi(z_{-U+1},\dots,z_U) = \sup_{(j_1,j_2)\in Q} \Theta(z_{-U+1},\dots,z_U,j_1,j_2). \tag{2.5.235}$$

In order to show continuity of Ψ, we first want to show that Θ is continuous. Clearly it is continuous with respect to any z_i when fixing j_1 and j_2. Fix $z^{(0)} \in \mathbb{R}^{2U}$ and $(j_1,j_2) \in Q$. Let $\epsilon_1 > 0$. We can choose $\delta > 0$ such that for any $z \in \mathbb{R}^{2U}$ we have

$$\|z - z^{(0)}\| < \delta \implies |\Theta(z,j_1,j_2) - \Theta(z^{(0)},j_1,j_2)| < \epsilon_1. \tag{2.5.236}$$

Take $\delta_0 = \min\{1,\delta\}$. Suppose $z \in \mathbb{R}^{2U}$ and $i_1, i_2 \in Q$ such that

$$\left\|\begin{pmatrix} z \\ i_1 \\ i_2 \end{pmatrix} - \begin{pmatrix} z^{(0)} \\ j_1 \\ j_2 \end{pmatrix}\right\| < \delta_0. \tag{2.5.237}$$

Then

$$\left\|\begin{pmatrix} i_1 \\ i_2 \end{pmatrix} - \begin{pmatrix} j_1 \\ j_2 \end{pmatrix}\right\| < 1, \tag{2.5.238}$$

which implies that $i_1 = j_1$ and $j_1 = j_2$. Therefore

$$\|z - z^{(0)}\| = \left\|\begin{pmatrix} z \\ i_1 \\ i_2 \end{pmatrix} - \begin{pmatrix} z^{(0)} \\ j_1 \\ j_2 \end{pmatrix}\right\| < \delta_0 \le \delta \tag{2.5.239}$$

and hence

$$|\Theta(z, i_1, i_2) - \Theta(z^{(0)}, j_1, j_2)| = |\Theta(z, j_1, j_2) - \Theta(z^{(0)}, j_1, j_2)| < \epsilon_1, \tag{2.5.240}$$

from which we can conclude that Θ is continuous on its domain. We will use this to show that Ψ is continuous. Fix $z^{(0)} \in \mathbb{R}^{2U}$. Let $\epsilon_2 > 0$. By continuity of Θ there exists $\delta > 0$ such that for any $z \in \mathbb{R}^{2U}$ and $i, j \in Q$ we have

$$\left\| \begin{pmatrix} z^{(0)} \\ i \end{pmatrix} - \begin{pmatrix} z \\ j \end{pmatrix} \right\| < \delta \implies |\Theta(z^{(0)}, i) - \Theta(z, j)| < \frac{\epsilon_2}{2}. \tag{2.5.241}$$

Then for any $z \in \mathbb{R}^{2U}$ with $\|z - z^{(0)}\| < \delta$ and any $i \in Q$ we have

$$\left\| \begin{pmatrix} z^{(0)} \\ i \end{pmatrix} - \begin{pmatrix} z \\ i \end{pmatrix} \right\| = \|z - z^{(0)}\| < \delta \tag{2.5.242}$$

and therefore

$$|\Theta(z^{(0)}, i) - \Theta(z, i)| < \epsilon_2. \tag{2.5.243}$$

Hence for any $i \in Q$ we have

$$\Theta(z, i) \leq \Theta(z^{(0)}, i) + \epsilon_2 \leq \sup_{j \in Q} \Theta(z^{(0)}, j) + \epsilon_2 = \Psi(z^{(0)}) + \epsilon_2 \tag{2.5.244}$$

and

$$\Theta(z^{(0)}, i) \leq \Theta(z, i) + \epsilon_2 \leq \sup_{j \in Q} \Theta(z, j) + \epsilon_2 = \Psi(z) + \epsilon_2 \tag{2.5.245}$$

Since Q is compact, $\Theta(z^{(0)}, i)$ and $\Theta(z, i)$ attain their supremum. Fix $i, j \in Q$ such that

$$\Theta(z^{(0)}, i) = \Psi(z^{(0)}) \quad \text{and} \quad \Theta(z, j) = \Psi(z). \tag{2.5.246}$$

Then

$$\Psi(z^{(0)}) = \Theta(z^{(0)}, i) \leq \Psi(z) + \epsilon_2 \tag{2.5.247}$$

and

$$\Psi(z) = \Theta(z, j) \leq \Psi(z^{(0)}) + \epsilon_2. \tag{2.5.248}$$

Thus

$$|\Psi(z^{(0)}) - \Psi(z)| \leq \epsilon_2, \tag{2.5.249}$$

from which we can conclude that Ψ is a continuous function. Since

$$(y_{-U+1}^{(n)}, \ldots, y_U^{(n)}) \xrightarrow{d} (y_{-U+1}, \ldots, y_U), \tag{2.5.250}$$

and Ψ is continuous, we can apply the continuous mapping theorem to find that

$$Z_n = \Psi(y_{-U+1}^{(n)}, \ldots, y_U^{(n)}) \xrightarrow{d} \Psi(y_{-U+1}, \ldots, y_U). \tag{2.5.251}$$

■

Chapter 3

Convergence speed in 2D model

The random variable we consider in this chapter should have different constant densities in two parts of the unit square. We split the unit interval into two parts. We do this with the parameters $\lambda, \mu \in (0, 1)$. These parameters refer to points on the bottom and top edge of the square respectively. The square is split into two parts by the straight line connecting $(\lambda, 0)$ with $(\mu, 1)$. This creates areas to the left and to the right of the line, which are denoted L and R respectively. The random variable lies in these areas with probabilities θ and $1 - \theta$ respectively. Within these areas, the random variable is uniformly distributed. This gives us a joint probability distribution function of the form

$$f(x, y) = \begin{cases} \beta, & \text{if } (x, y) \in L, \\ \gamma, & \text{if } (x, y) \in R, \\ 0, & \text{otherwise,} \end{cases} \tag{3.0.1}$$

where

$$\iint_L f(x, y) dx dy = \beta \cdot \iint_L 1 dx dy = \theta, \tag{3.0.2}$$

and

$$\iint_R f(x, y) dx dy = \left(1 - \iint_L 1 dx dy\right) \gamma = 1 - \theta. \tag{3.0.3}$$

In order to be able to be precise, we introduce notation that takes the parameters λ and μ into account. In order to do this, we first notice that the line segment connecting $(\lambda, 0)$ with $(\mu, 1)$ is given by

$$\{(x, y) \in [0, 1]^2 : x = (\mu - \lambda) y + \lambda\}. \tag{3.0.4}$$

Therefore the left and right side of the line segment are given by

$$\{(x, y) \in [0, 1]^2 : x \leq (\mu - \lambda) y + \lambda\}, \tag{3.0.5}$$

and

$$\{(x, y) \in [0, 1]^2 : x > (\mu - \lambda) y + \lambda\}, \tag{3.0.6}$$

respectively.

The left and ride side areas of the unit square, based on the values for the parameters λ and μ are

$$L(\lambda, \mu) := \{(x, y) \in [0, 1]^2 : x \leq (\mu - \lambda)y + \lambda\}, \tag{3.0.7}$$

and

$$R(\lambda, \mu) := \{(x, y) \in [0, 1]^2 : x > (\mu - \lambda)y + \lambda\}, \tag{3.0.8}$$

respectively.

In the one dimensional case, we worked with a parameter α. This was the parameter that we wanted to estimate with the MLE, as it defined the location of the boundary. In the two dimensional case we still define a value α. This time it is the area of the subset of $[0, 1]^2$ to the left side of the boundary that divides the square. In this way it will sometimes show up in the equations in the same way as it did in the one dimensional case. However, it does not define the location of the boundary, so this time estimation of α will not be of prime interest. We denote the area of $L(\lambda, \mu)$ by

$$\alpha(\lambda, \mu) := \iint_{L(\lambda,\mu)} 1 dx dy. \tag{3.0.9}$$

Since $L(\lambda, \mu)$ is a trapezoid, we find that $\alpha(\lambda, \mu) = \frac{1}{2}(\lambda + \mu)$. Notice that the area of $R(\lambda, \mu)$ is given by $1 - \alpha(\lambda, \mu)$. This naturally leads us to define β and γ as functions of λ, μ and θ. The functions $\beta, \gamma : (0, 1)^3 \to (0, \infty)$ are defined by

$$\beta(\lambda, \mu, \theta) := \frac{\theta}{\alpha(\lambda, \mu)} \quad \text{and} \quad \gamma(\lambda, \mu, \theta) := \frac{1 - \theta}{1 - \alpha(\lambda, \mu)}. \tag{3.0.10}$$

Now these are defined such that for all $\lambda, \mu, \theta \in (0, 1)$ replacing β and γ in Equation (3.0.1) with their respective functions from Equation (3.0.10) yields a proper probability density function. For fixed parameters $\lambda, \mu, \theta \in (0, 1)$ the probability density function $f : \mathbb{R}^2 \to [0, \infty)$ is defined by

$$f(x, y, \lambda, \mu, \theta) := \begin{cases} \beta(\lambda, \mu, \theta), & \text{if } (x, y) \in L(\lambda, \mu), \\ \gamma(\lambda, \mu, \theta), & \text{if } (x, y) \in R(\lambda, \mu), \\ 0, & \text{otherwise.} \end{cases} \tag{3.0.11}$$

We are interested in maximum likelihood estimation. Therefore we need to sample from the distribution. To this end we fix $\lambda_0, \mu_0, \theta_0 \in (0, 1)$ and let $X_1, X_2, \ldots$ be i.i.d. distributed on $[0, 1]^2$ with joint density $f(x, y, \lambda_0, \mu_0, \theta_0)$. These samples will be used by the estimator. Since we will often work with the density $f(x, y, \lambda_0, \mu_0, \theta_0)$ of the random variables X_i we introduce the some notation.

The values $\alpha_0 \in (0, 1)$ and $\beta_0, \gamma_0 \in \mathbb{R}^+$ are defined by

$$\alpha_0 := \alpha(\lambda_0, \mu_0), \qquad \beta_0 := \beta(\lambda_0, \mu_0, \theta_0), \qquad \text{and} \qquad \gamma_0 := \gamma(\lambda_0, \mu_0, \theta_0). \tag{3.0.12}$$

The areas $L_0, R_0 \subset [0, 1]^2$ are defined by

$$L_0 := L(\lambda_0, \mu_0), \qquad \text{and} \qquad R_0 := R(\lambda_0, \mu_0). \tag{3.0.13}$$

Assumption 3.1. *Since it is assumed that f is not the PDF for the uniform distribution on the unit square and hence the densities on the two sides of the boundary are different, we can also assume that $\alpha_0 \neq \theta_0$.*

Assumption 3.2. *Without loss of generality we assume that $\beta_0 > \gamma_0$.*

In this two dimensional case the joint CDF is not very useful. This is the case because we care more about the probability that observations lie on the left side of straight lines. This will be the analog to the CDF that we used in one dimension. Therefore we use the following notation.

Definition 3.1 (Cumulative line distribution function)**.** The *cumulative line distribution function* or CLDF corresponding to distribution $f(x, y, \lambda_0, \mu_0, \theta_0)$ is defined as

$$F_0(\lambda, \mu) := \mathbb{P}(X \in L(\lambda, \mu)), \tag{3.0.14}$$

where X is some random variable which is distributed with PDF $f(x, y, \lambda_0, \mu_0, \theta_0)$.

For the same reason, we want to define an analog to the eCDF that we used in one dimension. This will give the ratio of observations that lie in $L(\lambda, \mu)$ for any $\lambda, \mu \in (0, 1)$. For this we use the following notation.

Definition 3.2 (Empirical cumulative line distribution function)**.** For any $n \in \mathbb{N}$ the *Empirical cumulative line distribution function* eCLDF for $X_1, \ldots, X_n$ is defined as

$$F_n(\lambda, \mu) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L(\lambda, \mu)}. \tag{3.0.15}$$

3.1 Maximum likelihood estimators

The goal of this section are to find an expression for the MLE of θ_0 in terms of the MLEs of λ_0 and μ_0, and to find a convenient form for the MLEs of λ_0 and μ_0. To compute the MLEs, we look at the likelihood

$$\mathcal{L}_n(\lambda, \mu, \theta) = \prod_{i=1}^{n} f(X_i, \lambda, \mu, \theta) \tag{3.1.1}$$

$$= \prod_{X_i \in L(\lambda, \mu)} \beta(\lambda, \mu, \theta) \prod_{X_i \in R(\lambda, \mu)} \gamma(\lambda, \mu, \theta) \tag{3.1.2}$$

$$= (\beta(\lambda, \mu, \theta))^{nF_n(\lambda, \mu)} (\gamma(\lambda, \mu, \theta))^{n(1 - F_n(\lambda, \mu))}. \tag{3.1.3}$$

We notice that this expression is the same as the likelihood in the one-dimensional case, but with $F_n(\alpha)$ replaced by $F_n(\lambda, \mu)$ and α replaced by $\alpha(\lambda, \mu)$. This is the case since we have defined $\alpha(\lambda, \mu)$ and $F_n(\lambda, \mu)$ in such a way that they are the two-dimensional equivalents of $F_n(\alpha)$ and α. Notice that α in $F_n(\alpha)$ is not the same as the other α which is replaced by $\alpha(\lambda, \mu)$. We can find the MLEs using the same computations as in one dimension, leading us to the next definition.

Definition 3.3. The MLEs for λ_0, μ_0 and θ_0 are denoted by

$$(\hat{\lambda}_n, \hat{\mu}_n) := \underset{(x_1,x_2)\in(0,1)^2}{\arg\max} \Bigg(\log\left(\frac{F_n(x_1,x_2)}{\alpha(x_1,x_2)}\right) F_n(x_1,x_2) + \log\left(\frac{1-F_n(x_1,x_2)}{1-\alpha(x_1,x_2)}\right)(1-F_n(x_1,x_2)) \Bigg), \tag{3.1.4}$$

and

$$\hat{\theta}_n := F_n(\hat{\lambda}_n, \hat{\mu}_n). \tag{3.1.5}$$

The proof of the fact that these are indeed the MLEs is omitted here, as it is the exact same as in one dimension. For the sake of completeness, the proof is included in Appendix Section A.3.

3.2 Consistency of MLEs

The form of the function that the MLEs $(\hat{\lambda}_n, \hat{\mu}_n)$ is practically the same as in the one-dimensional case. This leads us to believe that these MLEs will also be consistent.

Proposition 3.1. *The MLE for* (λ_0, μ_0) *have the property that*

$$\|(\hat{\lambda} - \lambda_0, \hat{\mu} - \mu_0)\| = o_p(1). \tag{3.2.1}$$

We do not prove Conjecture 3.1, but we will discuss the potential pitfalls in the proof. To emulate the one-dimensional proof, the MLEs will be written as a function of eCLDFs. The first step will be to show that evaluating this function in F_0 yields (λ_0, μ_0). It seems that this can work in the same way as the one-dimensional proof. However, this time multiple partial derivatives have to be taken, because there are multiple variables. After that is done, there will be a Lemma stating that evaluating the function in a sequence of deterministic CLDFs under some conditions converges to (λ_0, μ_0). Overall it seems that this is not very different from the one-dimensional case. The fact that G_n is not a function of α, while the denominator is, can change some details though. After this lemma, it needs to be shown that F_n has the correct properties in order to apply the probabilistic version of the lemma. This time, the Glivenko-Cantelli theorem can not be used, as we are not working with eCDFs, but eCLDFs. For this part of the proof, a counterpart to the Glivenko-Cantelli theorem needs to be proven. For the rest of the proof, it will be necessary to manipulate some random variables in order to be able to work with a uniform random variable. In this case it will likely give us a uniform distribution on the unit square. The details will be a bit more involved when working with this distribution.

For the $\hat{\theta}$, we base the following conjecture on Conjecture 3.1.

Proposition 3.2. *The MLE for* θ_0 *has the property that*

$$\hat{\theta} - \theta_0 = o_p(1). \tag{3.2.2}$$

The proof of Conjecture 3.2 is entirely analogous to the one-dimensional case. It uses consistency of the other MLEs.

3.3 Quasi-maximum likelihood estimators

In the same way as in the one-dimensional case, we want to introduce a QMLE. This QMLE should be close enough to the MLE, but still have the convergence property that we want. We could try to look at the likelihood in the same way as in one dimension. Let us start by presenting the QMLEs.

The function that we want to use for defining the QMLE is denoted by

$$M_n(\lambda,\mu,\theta) := F_n(\lambda,\mu) - F_n(\lambda_0,\mu_0) + \frac{(\gamma_0-\beta_0)(\lambda-\lambda_0+\mu-\mu_0)\frac{1}{2}}{\log(\beta(\lambda_0,\mu_0,\theta)) - \log(\gamma(\lambda_0,\mu_0,\theta))}. \tag{3.3.1}$$

Based on Lemma 2.8 and Conjecture 3.1 we define the sequence $(a_n)_{n\in\mathbb{N}}$ of positive numbers with $\lim_{n\to\infty} a_n = 0$ such that

$$\|(\hat{\lambda}_n - \lambda_0, \hat{\mu}_n - \mu_0)\| = o_p(a_n), \tag{3.3.2}$$

as $n \to \infty$.

Definition 3.4. The QMLEs $\tilde{\lambda}$ and $\tilde{\mu}$ for λ_0 and μ_0 are defined as

$$(\tilde{\lambda}_n, \tilde{\mu}_n) := \underset{\|(\lambda-\lambda_0,\mu-\mu_0)\|\leq a_n}{\arg\max} M_n(\lambda,\mu,\theta_0). \tag{3.3.3}$$

The definition of M_n is almost the same as its one-dimensional equivalent. As this thesis does not contain the proofs for Sections 3.4 and 3.5, it is still possible that slight adjustments need to be done. Of course the function M_n only works well if we have a the same asymptotic properties of the log-likelihood as in the one-dimensional case. For this we state the following theorem, which will be proven later on in this section.

Theorem 3.3. *The log-likelihood has the property that*

$$\ell_n(\lambda,\mu,\theta) = \frac{1}{n}\sum_{i=1}^{n} \log(f(X_i,\lambda_0,\mu_0,\theta)) + S_n(\lambda,\mu,\theta), \tag{3.3.4}$$

where

$$\begin{aligned} S_n(\lambda,\mu,\theta) = &\left(\log\left(\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)}\right) + o(1)\right) M_n(\lambda,\mu,\theta) \\ &+ \frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(o(1)+o_p(1)) + T_p\left(\|(\lambda-\lambda_0,\mu-\mu_0)\|\right), \end{aligned} \tag{3.3.5}$$

where o and T_p are for $(\lambda,\mu)\to(\lambda_0,\mu_0)$ and o_p is for $n\to\infty$.

From Theorem 3.3 we can see that indeed $M_n(\lambda,\mu,\theta)$ is defined in a way such that the log-likelihood has similar properties as in one dimension. Notice that $\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0) = \alpha(\lambda,\mu)-\alpha_0$. This brings it even closer to the one-dimensional case. We will now briefly comment on the difference between the analysis that was done for Theorem 3.3 and the analysis in the one-dimensional case. Afterwards we will give the proof of the theorem.

The quantities $S_n^{(1)}, S_n^{(2)}$, and S_n are defined in the same way as in one dimension. Now $S_n^{(1)}$ summed over the points between the two lines created by (λ, μ) and (λ_0, μ_0). The difference here is that now the densities with respect to the two distributions are different in the two triangles that are created if the two lines cross. To compensate for this, we get an extra term, which turns out to have good asymptotic properties. In the end, this term combines with a term from $S_n^{(2)}$ to give us $T_p\left(\|(\lambda - \lambda_0, \mu - \mu_0)\|\right)$ as seen in Theorem 3.3. Hence this term is not a problem.

It is crucial that we want to have asymptotic terms with $\|(\lambda - \lambda_0, \mu - \mu_0)\|$ in them. Blindly following the steps from the one-dimensional case would not give these kinds of results. Instead they would give asymptotic terms with $\alpha(\lambda, \mu) - \alpha_0$ in them. This difference was very prominent in the analysis of $S_n^{(2)}$. At some point in this analysis, a function needs to be linearised. In the one-dimensional case, it was linearised with respect to α. This time, it has to be linearised with respect to λ and μ. To get to the correct asymptotic properties, we write the error term as a function multiplied by $\|(\lambda - \lambda_0, \mu - \mu_0)\|$. Except for some details, the rest is similar.

Now we will give the proof of Theorem 3.3.

Proof of Theorem 3.3. We will start by rewriting the log-likelihood. The likelihood is given by

$$\mathcal{L}_n(\lambda, \mu, \theta) = \prod_{i=1}^{n} f(X_i, \lambda, \mu, \theta). \tag{3.3.6}$$

Instead we can look at the log-likelihood divided by n, which is given by

$$\ell_n(\lambda, \mu, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log(f(X_i, \lambda, \mu, \theta)) \tag{3.3.7}$$

$$\begin{aligned} &= \frac{1}{n} \sum_{i=1}^{n} \log(f(X_i, \lambda_0, \mu_0, \theta)) \\ &+ \frac{1}{n} \sum_{i=1}^{n} \big(\log(f(X_i, \lambda, \mu, \theta) - \log(f(X_i, \lambda_0, \mu_0, \theta)\big). \end{aligned} \tag{3.3.8}$$

The first sum in Equation (3.3.8) does not depend on λ and μ. Hence it makes sense to ignore that part when constructing the QMLE for (λ_0, μ_0). The second part of Equation (3.3.8) is denoted by

$$S_n(\lambda, \mu, \theta) := \frac{1}{n} \sum_{i=1}^{n} \big(\log(f(X_i, \lambda, \mu, \theta) - \log(f(X_i, \lambda_0, \mu_0, \theta)\big). \tag{3.3.9}$$

We want to split this sum into two. The first sum will only include the observations that lie between the lines created by (λ, μ) and (λ_0, μ_0). The second sum will include all other observations. To be able to talk about the observations between the two lines and the observations outside of the two lines, we first introduce the following notation.

The area between the lines created by (λ, μ) and (λ_0, μ_0) is denoted by

$$B(\lambda, \mu) := \big(L(\lambda, \mu) \cap R_0\big) \cup \big(L_0 \cap R(\lambda, \mu)\big). \tag{3.3.10}$$

Now we can define the two separate sums using this notation.

The sum over the observations between the two lines is denoted by

$$S_n^{(1)}(\lambda, \mu, \theta) := \frac{1}{n} \sum_{X_i \in B(\lambda,\mu)} \big(\log(f(X_i, \lambda, \mu, \theta)) - \log(f(X_i, \lambda_0, \mu_0, \theta))\big), \tag{3.3.11}$$

and the sum over the other observations is denoted by

$$S_n^{(2)}(\lambda, \mu, \theta) := \frac{1}{n} \sum_{X_i \notin B(\lambda,\mu)} \big(\log(f(X_i, \lambda, \mu, \theta)) - \log(f(X_i, \lambda_0, \mu_0, \theta))\big). \tag{3.3.12}$$

Now we have $S_n(\lambda, \mu, \theta) = S_n^{(1)}(\lambda, \mu, \theta) + S_n^{(2)}(\lambda, \mu, \theta)$. We will rewrite $S_n^{(1)}$ and $S_n^{(2)}$ separately. We start by looking at $S_n^{(1)}$. First notice that for any observation $X_i \in B(\lambda, \mu)$

$$\begin{aligned} &\log(f(X_i, \lambda, \mu, \theta)) - \log(f(X_i, \lambda_0, \mu_0, \theta)) \\ &\qquad = \begin{cases} \log(\beta(\lambda, \mu, \theta)) - \log(\gamma(\lambda_0, \mu_0, \theta)), & \text{if } X_i \in L(\lambda, \mu) \cap R_0, \\ \log(\gamma(\lambda, \mu, \theta)) - \log(\beta(\lambda_0, \mu_0, \theta)), & \text{if } X_i \in L_0 \cap R(\lambda, \mu). \end{cases} \end{aligned} \tag{3.3.13}$$

Therefore

$$\begin{aligned} S_n^{(1)}(\lambda, \mu, \theta) = &\left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L(\lambda,\mu) \cap R_0} \right) \big(\log(\beta(\lambda, \mu, \theta)) - \log(\gamma(\lambda_0, \mu_0, \theta))\big) \\ &- \left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} \right) \big(\log(\beta(\lambda_0, \mu_0, \theta)) - \log(\gamma(\lambda, \mu, \theta))\big). \end{aligned} \tag{3.3.14}$$

Using the definition of the eCLDF from Definition 3.2, note that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L(\lambda,\mu) \cap R_0} - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} = F_n(\lambda, \mu) - F_n(\lambda_0, \mu_0). \tag{3.3.15}$$

We can use this to rewrite the expression for $S_n^{(1)}(\lambda, \mu, \theta)$ from Equation (3.3.14) into

$$S_n^{(1)}(\lambda, \mu, \theta) = (F_n(\lambda, \mu) - F_n(\lambda_0, \mu_0))\big(\log(\beta(\lambda_0, \mu_0, \theta)) - \log(\gamma(\lambda_0, \mu_0, \theta))\big) + \xi_n(\lambda, \mu, \theta), \tag{3.3.16}$$

where

$$\begin{aligned} \xi_n(\lambda, \mu, \theta) := &\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L(\lambda,\mu) \cap R_0} \big(\log(\beta(\lambda, \mu, \theta)) - \log(\beta(\lambda_0, \mu_0, \theta))\big) \\ &- \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} \big(\log(\gamma(\lambda_0, \mu_0, \theta)) - \log(\gamma(\lambda, \mu, \theta))\big). \end{aligned} \tag{3.3.17}$$

Unlike the one-dimensional equivalent of ξ_n, we now have a function that is not multiplied by $F_n(\lambda, \mu) - F_n(\lambda_0, \mu_0)$. We can rewrite it as a function that does include this expression, but it will then include another term to compensate for it. This term needs to have good asymptotic properties. By adding and subtracting a term, we can write

$$\begin{aligned}\xi_n(\lambda, \mu, \theta) &= \frac{1}{n}\sum_{i=1}^{n} \left(\mathbf{1}_{X_i \in L(\lambda,\mu)\cap R_0} - \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)}\right) \log \frac{\beta(\lambda, \mu, \theta)}{\beta(\lambda_0, \mu_0, \theta)} \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} \log \frac{\gamma(\lambda_0, \mu_0, \theta)\beta(\lambda_0, \mu_0, \theta)}{\gamma(\lambda, \mu, \theta)\beta(\lambda, \mu, \theta)}\end{aligned} \tag{3.3.18}$$

$$\begin{aligned}&= (F_n(\lambda, \mu) - F_n(\lambda_0, \mu_0)) \log \frac{\beta(\lambda, \mu, \theta)}{\beta(\lambda_0, \mu_0, \theta)} \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} \log \frac{\alpha(\lambda, \mu)(1-\alpha(\lambda, \mu))}{\alpha_0(1-\alpha_0)}.\end{aligned} \tag{3.3.19}$$

Recall the definition of T_p from Definition 2.1. We want to show that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} \log \frac{\alpha(\lambda, \mu)(1-\alpha(\lambda, \mu))}{\alpha_0(1-\alpha_0)} = T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.20}$$

for $(\lambda, \mu) \to (\lambda_0, \mu_0)$. To do this, it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda,\mu)} = T_p(1), \tag{3.3.21}$$

and

$$\log \frac{\alpha(\lambda, \mu)(1-\alpha(\lambda, \mu))}{\alpha_0(1-\alpha_0)} = O(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.22}$$

both for $(\lambda, \mu) \to (\lambda_0, \mu_0)$. We start by showing the former. To prove this statement, we let (η_n^1, η_n^2) be a consistent estimator of (λ_0, μ_0). Then

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\eta_n^1, \eta_n^2)} \sim \frac{1}{n}\text{Binom}\left(n, \iint_{L_0 \cap R(\eta_n^1, \eta_n^2)} \beta_0 dt\right). \tag{3.3.23}$$

Define

$$\omega_n := \iint_{L_0 \cap R(\eta_n^1, \eta_n^2)} \beta_0 dt. \tag{3.3.24}$$

By consistency of (η_n^1, η_n^2), we have

$$\lim_{n\to\infty} \omega_n = 0. \tag{3.3.25}$$

Let $\epsilon > 0$. We can choose $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\omega_n < \epsilon/2$. Then for $n \geq N$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\eta_n^1, \eta_n^2)} < \epsilon\right) \geq \mathbb{P}\left(\frac{1}{n}\text{Binom}(n, \epsilon/2) < \epsilon\right). \tag{3.3.26}$$

By the law of large numbers

$$\frac{1}{n}\text{Binom}(n, \epsilon/2) \xrightarrow{a.s.} \epsilon/2. \tag{3.3.27}$$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\text{Binom}(n, \epsilon/2) < \epsilon\right) = 1. \tag{3.3.28}$$

By the squeeze theorem we conclude that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\eta_n^1, \eta_n^2)} < \epsilon\right) = 1, \tag{3.3.29}$$

and hence

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \in L_0 \cap R(\lambda, \mu)} = T_p(1), \tag{3.3.30}$$

for $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Now we will prove that

$$\log \frac{\alpha(\lambda, \mu)(1-\alpha(\lambda, \mu))}{\alpha_0(1-\alpha_0)} = O(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.31}$$

for $(\lambda, \mu) \to (\lambda_0, \mu_0)$. To do this we analyse

$$\left|\frac{\log \frac{(\lambda+\mu)(2-\lambda-\mu)}{(\lambda_0+\mu_0)(2-\lambda_0-\mu_0)}}{\|(\lambda - \lambda_0, \mu - \mu_0)\|}\right|. \tag{3.3.32}$$

We will rewrite this expression by writing λ and μ as polar coordinates with center (λ_0, μ_0). This gives the substitution

$$\lambda = \lambda_0 + r\cos\phi, \qquad \text{and} \qquad \mu = \mu_0 + r\sin\phi. \tag{3.3.33}$$

Plugging this into Equation (3.3.32) yields

$$\frac{1}{r}\left|\log\left(1 + \frac{r(\cos\phi + \sin\phi)}{\lambda_0 + \mu_0}\right) + \log\left(1 - \frac{r(\cos\phi + \sin\phi)}{2 - \lambda_0 + \mu_0}\right)\right|. \tag{3.3.34}$$

We want to bound this for any fixed r. To do this we first apply the triangle inequality. Using that for any $x > 0$ we have $|\log(1-x)| \geq |\log(1+x)|$ and that for any ϕ we have $\cos\phi + \sin\phi \in [-\sqrt{2}, \sqrt{2}]$, we find that

$$\left|\log\left(1 + \frac{r(\cos\phi + \sin\phi)}{\lambda_0 + \mu_0}\right)\right| \leq \left|\log\left(1 - r\frac{\sqrt{2}}{\lambda_0 + \mu_0}\right)\right|, \tag{3.3.35}$$

and

$$\left|\log\left(1 - \frac{r(\cos\phi + \sin\phi)}{2 - \lambda_0 + \mu_0}\right)\right| \leq \left|\log\left(1 - r\frac{\sqrt{2}}{2 - \lambda_0 + \mu_0}\right)\right|. \tag{3.3.36}$$

Therefore

$$\left|\frac{\log \frac{(\lambda+\mu)(2-\lambda-\mu)}{(\lambda_0+\mu_0)(2-\lambda_0-\mu_0)}}{\|(\lambda-\lambda_0,\mu-\mu_0)\|}\right| \leq \frac{1}{r}\left(\left|\log\left(1-r\frac{\sqrt{2}}{\lambda_0+\mu_0}\right)\right| + \left|\log\left(1-r\frac{\sqrt{2}}{2-\lambda_0+\mu_0}\right)\right|\right). \tag{3.3.37}$$

We want to show that this bound is increasing in r when r is positive, since that would mean that a bound for a fixed r also bounds everything within a distance of r from (λ_0, μ_0). Let $a > 0$ be fixed. Then $\log(1 - ax)$ is negative for positive x. Thus $\log(1 - ax)/x$ is also negative for $x > 0$. We need to show that $\log(1 - ax)/x$ is decreasing, because this would imply that $|\log(1 - ax)/x|$ is increasing for $x > 0$. To do this, we first find that the derivative is given by

$$\frac{-ax-(1-ax)\log(1-ax)}{(1-ax)x^2}. \tag{3.3.38}$$

Since we only consider $x \in (0, 1/a)$, we find that $(1 - ax)x^2 > 0$. We need to show that $-ax - (1 - ax)\log(1 - ax) \leq 0$ for $x > 0$. For this we use that $\log y \geq 1 - 1/y$ for any $y > 0$. Applying this to $1 - ax$ we find that $\log(1 - ax) \geq 1 - 1/(1 - ax)$. But then also

$$-(1-ax)\log(1-ax) \leq -(1-ax)\left(1-\frac{1}{1-ax}\right) = ax. \tag{3.3.39}$$

Thus $-ax - (1 - ax)\log(1 - ax) \leq -ax + ax = 0$, from which we can conclude that the derivative is non-positive and thus $|\log(1 - ax)/x|$ is increasing for $x > 0$. By taking the correct values for a we find that

$$\frac{1}{r}\left(\left|\log\left(1-r\frac{\sqrt{2}}{\lambda_0+\mu_0}\right)\right| + \left|\log\left(1-r\frac{\sqrt{2}}{2-\lambda_0+\mu_0}\right)\right|\right), \tag{3.3.40}$$

is increasing in r for $r > 0$. Now we can fix some $r > 0$ such that

$$1-r\frac{\sqrt{2}}{\lambda_0+\mu_0} > 0, \quad \text{and} \quad 1-r\frac{\sqrt{2}}{2-\lambda_0+\mu_0} > 0. \tag{3.3.41}$$

Then for all (λ, μ) such that $\|(\lambda - \lambda_0, \mu - \mu_0)\| \leq r$, we have

$$\left|\frac{\log \frac{(\lambda+\mu)(2-\lambda-\mu)}{(\lambda_0+\mu_0)(2-\lambda_0-\mu_0)}}{\|(\lambda-\lambda_0,\mu-\mu_0)\|}\right| \leq \frac{1}{r}\left(\left|\log\left(1-r\frac{\sqrt{2}}{\lambda_0+\mu_0}\right)\right| + \left|\log\left(1-r\frac{\sqrt{2}}{2-\lambda_0+\mu_0}\right)\right|\right), \tag{3.3.42}$$

where this time r is the fixed value that we chose. We conclude that indeed

$$\log\frac{\alpha(\lambda,\mu)(1-\alpha(\lambda,\mu))}{\alpha_0(1-\alpha_0)} = O(\|(\lambda-\lambda_0,\mu-\mu_0)\|), \tag{3.3.43}$$

for $(\lambda, \mu) \to (\lambda_0, \mu_0)$ and therefore

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{X_i\in L_0\cap R(\lambda,\mu)}\log\frac{\alpha(\lambda,\mu)(1-\alpha(\lambda,\mu))}{\alpha_0(1-\alpha_0)} = T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|), \tag{3.3.44}$$

for $(\lambda,\mu)\to(\lambda_0,\mu_0)$. Hence

$$S_n^{(1)}(\lambda,\mu,\theta) = (F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\left(\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)}+\log\frac{\beta(\lambda,\mu,\theta)}{\beta(\lambda_0,\mu_0,\theta)}\right) + T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|), \tag{3.3.45}$$

for $(\lambda,\mu)\to(\lambda_0,\mu_0)$. Define

$$\chi(\lambda,\mu) := \log\frac{\beta(\lambda,\mu,\theta)}{\beta(\lambda_0,\mu_0,\theta)} = \log\frac{\alpha_0}{\alpha(\lambda,\mu)}. \tag{3.3.46}$$

Notice that $\chi(\lambda,\mu)=o(1)$ as $(\lambda,\mu)\to(\lambda_0,\mu_0)$. Then

$$\begin{aligned} S_n^{(1)}(\lambda,\mu,\theta) &= (F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} + T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|)\\ &+ \chi(\lambda,\mu)\left(F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0)+\frac{(\gamma_0-\beta_0)(\lambda-\lambda_0+\mu-\mu_0)\frac{1}{2}}{\log(\beta(\lambda_0,\mu_0,\theta))-\log(\gamma(\lambda_0,\mu_0,\theta))}\right)\\ &- \chi(\lambda,\mu)\frac{(\gamma_0-\beta_0)(\lambda-\lambda_0+\mu-\mu_0)\frac{1}{2}}{\log(\beta(\lambda_0,\mu_0,\theta))-\log(\gamma(\lambda_0,\mu_0,\theta))} \end{aligned} \tag{3.3.47}$$

$$\begin{aligned} &= (F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} + T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|)\\ &- \chi(\lambda,\mu)\frac{(\gamma_0-\beta_0)(\lambda-\lambda_0+\mu-\mu_0)\frac{1}{2}}{\log(\beta(\lambda_0,\mu_0,\theta))-\log(\gamma(\lambda_0,\mu_0,\theta))} + \chi(\lambda,\mu)M_n(\lambda,\mu,\theta) \end{aligned} \tag{3.3.48}$$

$$\begin{aligned} &= (F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} + T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|)\\ &+ \frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)o(1) + o(1)M_n(\lambda,\mu,\theta), \end{aligned} \tag{3.3.49}$$

where o and T_p are for $(\lambda,\mu)\to(\lambda_0,\mu_0)$.

We now turn ourselves towards $S_n^{(2)}(\lambda,\mu,\theta)$. Note that for all $X_i\notin B(\lambda,\mu)$ we have

$$\begin{aligned} &\log(f(X_i,\lambda,\mu,\theta)) - \log(f(X_i,\lambda_0,\mu_0,\theta))\\ &\quad= \begin{cases}\log\beta(\lambda,\mu,\theta)-\log\beta(\lambda_0,\mu_0,\theta), & \text{if } X_i\in L(\lambda,\mu)\cap L_0,\\ \log\gamma(\lambda,\mu,\theta)-\log\gamma(\lambda_0,\mu_0,\theta), & \text{if } X_i\in R(\lambda,\mu)\cap R_0\end{cases} \qquad (3.3.50)\\ &\quad= \begin{cases}\log\frac{\alpha_0}{\alpha(\lambda,\mu)}, & \text{if } X_i\in L(\lambda,\mu)\cap L_0,\\ \log\frac{1-\alpha_0}{1-\alpha(\lambda,\mu)}, & \text{if } X_i\in R(\lambda,\mu)\cap R_0.\end{cases} \qquad (3.3.51) \end{aligned}$$

Notice that this value does not depend on θ. We can linearise this with respect to λ and μ around the point $(\lambda,\mu)=(\lambda_0,\mu_0)$. Define

$$b(x,\lambda,\mu) := \begin{cases}\log\frac{\alpha_0}{\alpha(\lambda,\mu)}, & \text{if } x\in L(\lambda,\mu)\cap L_0,\\ \log\frac{1-\alpha_0}{1-\alpha(\lambda,\mu)}, & \text{if } x\in R(\lambda,\mu)\cap R_0.\end{cases} \tag{3.3.52}$$

To linearise this around (λ_0, μ_0), we write it as

$$\begin{aligned} b(x, \lambda, \mu) = b(x, \lambda_0, \mu_0) &+ \frac{\partial b(x, \lambda_0, \mu_0)}{\partial \lambda}(\lambda - \lambda_0) \\ &+ \frac{\partial b(x, \lambda_0, \mu_0)}{\partial \mu}(\mu - \mu_0) + h(x, \lambda, \mu)\|(\lambda - \lambda_0, \mu - \mu_0)\|, \end{aligned} \tag{3.3.53}$$

for some function $h(x, \lambda, \mu)$. We can clearly see that for any x we have

$$b(x, \lambda_0, \mu_0) = 0. \tag{3.3.54}$$

Therefore

$$h(x, \lambda, \mu) = \frac{b(x, \lambda, \mu) - \frac{\partial b(x,\lambda_0,\mu_0)}{\partial \lambda}(\lambda - \lambda_0) - \frac{\partial b(x,\lambda_0,\mu_0)}{\partial \mu}(\mu - \mu_0)}{\|(\lambda - \lambda_0, \mu - \mu_0)\|}. \tag{3.3.55}$$

By using the chain rule we find that the partial derivative of b with respect to λ is given by

$$\frac{\partial b(x, \lambda, \mu)}{\partial \lambda} = \begin{cases} -\frac{1}{\alpha(\lambda,\mu)} \cdot \frac{1}{2}, & \text{if } x \in L(\lambda, \mu) \cap L_0, \\ \frac{1}{1-\alpha(\lambda,\mu)} \cdot \frac{1}{2}, & \text{if } x \in R(\lambda, \mu) \cap R_0. \end{cases} \tag{3.3.56}$$

The partial derivative of b with respect to μ is the same. Let

$$d(x) := \frac{\partial b(x, \lambda_0, \mu_0)}{\partial \lambda} = \begin{cases} -\frac{1}{\alpha_0} \cdot \frac{1}{2}, & \text{if } x \in L_0, \\ \frac{1}{1-\alpha_0} \cdot \frac{1}{2}, & \text{if } x \in R_0. \end{cases} \tag{3.3.57}$$

Then

$$h(x, \lambda, \mu) = \frac{b(x, \lambda, \mu) - d(x)(\lambda - \lambda_0 + \mu - \mu_0)}{\|(\lambda - \lambda_0, \mu - \mu_0)\|}. \tag{3.3.58}$$

Define

$$h_1(\lambda, \mu) := \frac{\log \frac{\alpha_0}{\alpha(\lambda,\mu)} + \frac{1}{2\alpha_0}(\lambda - \lambda_0 + \mu - \mu_0)}{\sqrt{(\lambda - \lambda_0)^2 + (\mu - \mu_0)^2}}, \tag{3.3.59}$$

and

$$h_2(\lambda, \mu) := \frac{\log \frac{1-\alpha_0}{1-\alpha(\lambda,\mu)} + \frac{1}{2(1-\alpha_0)}(\lambda - \lambda_0 + \mu - \mu_0)}{\sqrt{(\lambda - \lambda_0)^2 + (\mu - \mu_0)^2}}. \tag{3.3.60}$$

Then for $x \in L(\lambda, \mu) \cap L_0$ we have

$$h(x, \lambda, \mu) = h_1(\lambda, \mu) \tag{3.3.61}$$

and for $x \in R(\lambda, \mu) \cap R_0$ we have

$$h(x, \lambda, \mu) = h_2(\lambda, \mu). \tag{3.3.62}$$

By using b in the definition of $S_n^{(2)}(\lambda,\mu,\theta)$, we find that

$$S_n^{(2)}(\lambda,\mu,\theta) = \frac{1}{n}\sum_{X_i\notin B(\lambda,\mu)} b(X_i,\lambda,\mu) \tag{3.3.63}$$

$$= \frac{1}{n}\sum_{X_i\notin B(\lambda,\mu)} d(X_i)(\lambda-\lambda_0+\mu-\mu_0)$$

$$+ \frac{1}{n}\sum_{X_i\in L(\lambda,\mu)\cap L_0} h_1(\lambda,\mu)\|(\lambda-\lambda_0,\mu-\mu_0)\| \tag{3.3.64}$$

$$+ \frac{1}{n}\sum_{X_i\in R(\lambda,\mu)\cap R_0} h_2(\lambda,\mu)\|(\lambda-\lambda_0,\mu-\mu_0)\|$$

$$= \frac{1}{n}\sum_{X_i\notin B(\lambda,\mu)} d(X_i)(\lambda-\lambda_0+\mu-\mu_0)$$

$$+ \|(\lambda-\lambda_0,\mu-\mu_0)\|\left(h_1(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in L(\lambda,\mu)\cap L_0} 1\right)\right. \tag{3.3.65}$$

$$\left. + h_2(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in R(\lambda,\mu)\cap R_0} 1\right)\right).$$

We want to show that

$$\|(\lambda-\lambda_0,\mu-\mu_0)\|\left(h_1(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in L(\lambda,\mu)\cap L_0} 1\right) + h_2(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in R(\lambda,\mu)\cap R_0} 1\right)\right), \tag{3.3.66}$$

is $T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|)$ as $(\lambda,\mu)\to(\lambda_0,\mu_0)$. It suffices to show that

$$\left(h_1(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in L(\lambda,\mu)\cap L_0} 1\right) + h_2(\lambda,\mu)\left(\frac{1}{n}\sum_{X_i\in R(\lambda,\mu)\cap R_0} 1\right)\right) = T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|), \tag{3.3.67}$$

as $(\lambda,\mu)\to(\lambda_0,\mu_0)$. The sums

$$\left(\frac{1}{n}\sum_{X_i\in L(\lambda,\mu)\cap L_0} 1\right), \qquad \text{and} \qquad \left(\frac{1}{n}\sum_{X_i\in R(\lambda,\mu)\cap R_0} 1\right), \tag{3.3.68}$$

are bounded by 1 for any λ and μ and $n\in\mathbb{N}$. Thus evaluating them in consistent estimators results in sequences that are $O_p(1)$ as $n\to\infty$. Hence it suffices to show that $h_1(\lambda,\mu)$ and $h_2(\lambda,\mu)$ are $o(1)$ as $(\lambda,\mu)\to(\lambda_0,\mu_0)$. We start by investigating h_1. To solve the limit we will use polar coordinates around (λ_0, μ_0) and use the squeeze theorem. First we write

$$h_1(\lambda, \mu) = \frac{\log \frac{\lambda_0+\mu_0}{\lambda+\mu} + \frac{1}{\lambda_0+\mu_0}(\lambda - \lambda_0 + \mu - \mu_0)}{\sqrt{(\lambda-\lambda_0)^2 + (\mu-\mu_0)^2}} \tag{3.3.69}$$

$$= \frac{\log \frac{\lambda_0+\mu_0}{\lambda+\mu} + \frac{\lambda+\mu}{\lambda_0+\mu_0} - 1}{\sqrt{(\lambda-\lambda_0)^2 + (\mu-\mu_0)^2}}. \tag{3.3.70}$$

By using the substitution

$$\lambda = \lambda_0 + r\cos\phi, \qquad \text{and} \qquad \mu = \mu_0 + r\sin\phi, \tag{3.3.71}$$

we define

$$h_3(r, \phi) := h_1(\lambda_0 + r\cos\phi, \mu_0 + r\sin\phi). \tag{3.3.72}$$

Then

$$h_3(r,\phi) = \frac{\log \frac{\lambda_0+\mu_0}{\lambda_0+\mu_0+r(\sin\phi+\cos\phi)} + \frac{\lambda_0+\mu_0+r(\sin\phi+\cos\phi)}{\lambda_0+\mu_0} - 1}{r} \tag{3.3.73}$$

$$= \frac{1}{r}\log\frac{\lambda_0+\mu_0}{\lambda_0+\mu_0+r(\sin\phi+\cos\phi)} + \frac{\sin\phi+\cos\phi}{\lambda_0+\mu_0} \tag{3.3.74}$$

$$= \frac{-1}{r}\log\left(1 + \frac{r(\sin\phi+\cos\phi)}{\lambda_0+\mu_0}\right) + \frac{\sin\phi+\cos\phi}{\lambda_0+\mu_0}. \tag{3.3.75}$$

For convenience we define

$$\zeta(\phi) := \frac{\sin\phi+\cos\phi}{\lambda_0+\mu_0}. \tag{3.3.76}$$

Then

$$h_3(r,\phi) = \frac{-1}{r}\log\left(1 + r\zeta(\phi)\right) + \zeta(\phi). \tag{3.3.77}$$

Since we want to apply the squeeze theorem, we need to find a lower and an upper bound for $h_3(r, \phi)$ that only rely on r. This will allow us to take the limit of r going to 0 in the lower and upper bound, which will tell us about the limit of $h_1(\lambda, \mu)$. We start by finding a lower bound. Notice that we only need to consider $r > 0$. It is known that for any $y \in \mathbb{R}$ we have $\log(1 + y) \leq y$. Hence for any r and ϕ we have

$$h_3(r,\phi) = \frac{-1}{r}\log\left(1 + r\zeta(\phi)\right) + \zeta(\phi) \geq -\zeta(\phi) + \zeta(\phi) = 0. \tag{3.3.78}$$

For the upper bound, we first notice that we only have to consider r and ϕ such that $1 + r\zeta(\phi) > 0$. Since $\zeta(\phi)$ is 2π-periodic, h_3 is also 2π-periodic in ϕ. For r small enough, it is continuous. Hence it is a bounded function. The partial derivative with respect to ϕ of $h_3(r, \phi)$ is given by

$$\left(1 - \frac{1}{1 + r\zeta(\phi)}\right)\zeta'(\phi), \tag{3.3.79}$$

where

$$\zeta'(\phi) = \frac{\cos\phi - \sin(\phi)}{\lambda_0 + \mu_0}. \tag{3.3.80}$$

Hence the partial derivative of h_3 with respect to ϕ is zero if $\zeta'(\phi) = 0$ or $1 + r\zeta(\phi) = 1$. The latter happens if and only if $\zeta(\phi) = 0$. For such ϕ, the function h_3 is zero. Since the function is non-negative with a derivative which is not constantly zero, these can not be the maximum. Instead it happens when $\zeta'(\phi) = 0$, which is the case if and only if $\phi = \pi/4 + k \cdot \pi$ for some $k \in \mathbb{N}$. Since the function is 2π-periodic, we only need to consider the maxima on $[0, 2\pi]$. It is either $\pi/4$ or $5\pi/4$. We do not need to compute which of these it is, as we now know that $h_3(r, \phi)$ is bounded by $\max\{h_3(r, \pi/4), h_3(r, 5\pi/4)\}$ and we can show that both $h_3(r, \pi/4)$ and $h_3(r, 5\pi/4)$ converge to zero if we take r to zero. To show this, we fix $\phi_0 \in [0, 2\pi]$ and define $\zeta_0 := \zeta(\phi_0)$. We also define $h_4(r) := h_3(r, \phi_0)$. Then

$$h_4(r) = \frac{-1}{r}\log(1 + r\zeta_0) + \zeta_0 = \frac{r\zeta_0 - \log(1 + r\zeta_0)}{r}. \tag{3.3.81}$$

We can see that the limit of r going to zero of both the numerator and the denominator is zero. Therefore we can apply l'Hôpital's rule to find that

$$\lim_{r \to 0} h_4(r) = \lim_{r \to 0}\left(\zeta_0 - \frac{\zeta_0}{1 + r\zeta_0}\right) = 0. \tag{3.3.82}$$

Since this applies for all ϕ_0, we find that the upper bound for $h_3(r, \phi)$ converges to zero. Notice that we did have to do the above analysis, as the fact that this limit holds pointwise does not mean that it holds for the maximum. We did need to find a finite amount of candidates for the maximum. We can now apply the squeeze theorem to find that

$$\lim_{r \to 0} h_4(r, \phi) = 0, \tag{3.3.83}$$

and therefore

$$\lim_{(\lambda,\mu) \to (\lambda_0,\mu_0)} h_1(\lambda, \mu) = 0. \tag{3.3.84}$$

We conclude that $h_1(\lambda, \mu) = o(1)$ as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. In a similar way it can be shown that $h_2(\lambda, \mu) = o(1)$ as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Therefore indeed

$$\|(\lambda - \lambda_0, \mu - \mu_0)\| \left(h_1(\lambda, \mu) \left(\frac{1}{n} \sum_{X_i \in L(\lambda,\mu) \cap L_0} 1 \right) + h_2(\lambda, \mu) \left(\frac{1}{n} \sum_{X_i \in R(\lambda,\mu) \cap R_0} 1 \right) \right), \tag{3.3.85}$$

is $T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|)$ as $(\lambda, \mu) \to (\lambda_0, \mu_0)$ and thus

$$S_n^{(2)}(\lambda, \mu, \theta) = \frac{1}{n} \sum_{X_i \notin B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) + T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.86}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$

Now we want to rewrite the sum that we see in the equation above. To do this we rewrite it as the sum over all observations and subtract the observations in $B(\lambda, \mu)$. This gives us

$$\begin{aligned}\frac{1}{n}\sum_{X_i \notin B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) = &\frac{1}{n}\sum_{i=1}^{n} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) \\ &- \frac{1}{n}\sum_{X_i \in B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0).\end{aligned} \tag{3.3.87}$$

We want to show that

$$\frac{1}{n}\sum_{X_i \in B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) = T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.88}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Recall that

$$d(x) = \begin{cases} -\frac{1}{\alpha_0} \cdot \frac{1}{2}, & \text{if } x \in L_0, \\ \frac{1}{1-\alpha_0} \cdot \frac{1}{2}, & \text{if } x \in R_0. \end{cases} \tag{3.3.89}$$

Therefore

$$\begin{aligned}\frac{1}{n}\sum_{X_i \in B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) = &\left(\frac{1}{n}\sum_{X_i \in L_0 \cap R(\lambda,\mu)} 1\right) \frac{-(\lambda - \lambda_0 + \mu - \mu_0)}{2\alpha_0} \\ &+ \left(\frac{1}{n}\sum_{X_i \in R_0 \cap L(\lambda,\mu)} 1\right) \frac{\lambda - \lambda_0 + \mu - \mu_0}{2(1-\alpha_0)}.\end{aligned} \tag{3.3.90}$$

In Equation (3.3.30) it was shown that

$$\frac{1}{n}\sum_{X_i \in L_0 \cap R(\lambda,\mu)} 1 = T_p(1), \tag{3.3.91}$$

for $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Analogously it can be found that

$$\frac{1}{n}\sum_{X_i \in R_0 \cap L(\lambda,\mu)} 1 = T_p(1), \tag{3.3.92}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Hence it suffices to show that $\frac{-(\lambda-\lambda_0+\mu-\mu_0)}{2\alpha_0}$ and $\frac{\lambda-\lambda_0+\mu-\mu_0}{2(1-\alpha_0)}$ are $O(\|(\lambda - \lambda_0, \mu - \mu_0)\|)$ as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Dividing by $-2\alpha_0$ or $2(1 - \alpha_0)$ does not affect this behaviour. Thus it suffices to show that $\lambda - \lambda_0 + \mu - \mu_0$ is $O(\|(\lambda - \lambda_0, \mu - \mu_0)\|)$ as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Notice that

$$\frac{|\lambda - \lambda_0 + \mu - \mu_0|}{\|(\lambda - \lambda_0, \mu - \mu_0)\|} \leq \frac{|\lambda - \lambda_0| + |\mu - \mu_0|}{\|(\lambda - \lambda_0, \mu - \mu_0)\|} = \frac{\|(\lambda - \lambda_0, \mu - \mu_0)\|_1}{\|(\lambda - \lambda_0, \mu - \mu_0)\|_2}. \tag{3.3.93}$$

Since the 1-norm and the 2-norm on $\mathbb{R}^2$ are equivalent, this is bounded by some constant. Thus we find that indeed

$$\lambda - \lambda_0 + \mu - \mu_0 = O(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.94}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Therefore

$$\frac{1}{n} \sum_{X_i \in B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) = T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.95}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$.

Plugging this back into Equation (3.3.87) yields

$$\begin{aligned}\frac{1}{n} \sum_{X_i \notin B(\lambda,\mu)} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) = \frac{1}{n}\sum_{i=1}^{n} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) \\ + T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|),\end{aligned} \tag{3.3.96}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. Using this in Equation (3.3.86) we find that

$$\begin{aligned}S_n^{(2)}(\lambda, \mu, \theta) = \frac{1}{n}\sum_{i=1}^{n} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) + T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|) \\ + T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|)\end{aligned} \tag{3.3.97}$$

$$= \frac{1}{n}\sum_{i=1}^{n} d(X_i)(\lambda - \lambda_0 + \mu - \mu_0) + T_p(\|(\lambda - \lambda_0, \mu - \mu_0)\|), \tag{3.3.98}$$

as $(\lambda, \mu) \to (\lambda_0, \mu_0)$. By the law of large numbers we have

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i) \xrightarrow{p} \mathbb{E}\left[d(X)\right], \tag{3.3.99}$$

where X is a random variable which is distributed with probability density function $f(x, \lambda_0, \mu_0, \theta_0)$. This can also be written as

$$\frac{1}{n}\sum_{i=1}^{n} d(X_i) = \mathbb{E}\left[d(X_i)\right] + o_p(1), \tag{3.3.100}$$

for $n \to \infty$. Since

$$\mathbb{E}\left[d(X_i)\right] = \mathbb{P}(X \in L_0) \cdot \frac{-1}{2\alpha_0} + \mathbb{P}(X \in R_0) \cdot \frac{1}{2(1 - \alpha_0)} \tag{3.3.101}$$

$$= \frac{1 - \theta_0}{2(1 - \alpha_0)} - \frac{\theta_0}{2\alpha_0} \tag{3.3.102}$$

$$= \frac{1}{2}(\gamma_0 - \beta_0), \tag{3.3.103}$$

we find that

$$
\begin{aligned}
S_n^{(2)}(\lambda,\mu,\theta) = &\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(\gamma_0-\beta_0)+\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)o_p(1) \\
&+O_p(1)o(\|(\lambda-\lambda_0,\mu-\mu_0)\|),
\end{aligned}
\tag{3.3.104}
$$

where O_p, o_p are for $n\to\infty$ and o is for $(\lambda,\mu)\to(\lambda_0,\mu_0)$.

Now we want to combine $S_n^{(1)}(\lambda,\mu,\theta)$ and $S_n^{(2)}(\lambda,\mu,\theta)$ to get an expression for $S_n(\lambda,\mu,\theta)$. Recall that

$$
\begin{aligned}
S_n^{(1)}(\lambda,\mu,\theta) = &(F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} \\
&+T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|) \\
&+\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)o(1)+o(1)M_n(\lambda,\mu,\theta),
\end{aligned}
\tag{3.3.105}
$$

where T_p and o are for $(\lambda,\mu)\to(\lambda_0,\mu_0)$.

By using that $S_n(\lambda,\mu,\theta)=S_n^{(1)}(\lambda,\mu,\theta)+S_n^{(2)}(\lambda,\mu,\theta)$, we find that

$$
\begin{aligned}
S_n(\lambda,\mu,\theta) = &(F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} \\
&+T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|)+T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|) \\
&+\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(o(1)+o_p(1))+M_n(\lambda,\mu,\theta)o(1) \\
&+\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(\gamma_0-\beta_0),
\end{aligned}
\tag{3.3.106}
$$

where o and T_p are for $(\lambda,\mu)\to(\lambda_0,\mu_0)$ and o_p is for $n\to\infty$. Since

$$
\begin{aligned}
&\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(\gamma_0-\beta_0)+(F_n(\lambda,\mu)-F_n(\lambda_0,\mu_0))\log\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)} \\
&\quad=\log\left(\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)}\right)M_n(\lambda,\mu,\theta)
\end{aligned}
\tag{3.3.107}
$$

we can conclude that

$$
\begin{aligned}
S_n(\lambda,\mu,\theta) = &\left(\log\left(\frac{\beta(\lambda_0,\mu_0,\theta)}{\gamma(\lambda_0,\mu_0,\theta)}\right)+o(1)\right)M_n(\lambda,\mu,\theta) \\
&+\frac{1}{2}(\lambda-\lambda_0+\mu-\mu_0)(o(1)+o_p(1))+T_p(\|(\lambda-\lambda_0,\mu-\mu_0)\|),
\end{aligned}
\tag{3.3.108}
$$

where o and T_p are for $(\lambda,\mu)\to(\lambda_0,\mu_0)$ and o_p is for $n\to\infty$. Since most terms in this expression are asymptotically small and $\log(\beta(\lambda_0,\mu_0,\theta))-\log(\gamma(\lambda_0,\mu_0,\theta))+o(1)$ will be positive when θ is close enough to θ_0 and (λ,μ) is close enough to (λ_0,μ_0) it does seem to make sense to maximise $M(\lambda,\mu,\theta_0)$ in some region around (λ_0,μ_0). ■

3.4 Convergence speed of QMLEs

Because the form of M_n is roughly the same as in one dimension, the proof will likely be very similar. In the one-dimensional case, there needs to be a distinction on whether α is bigger or smaller than α_0. In two dimensions, this is not exactly possible, as there is the option of the lines crossing. This will give more problems. Another point is that in the one-dimensional case, we try to turn some problems into problems including uniform distributions on the unit interval. Again this means that we have to use uniform distributions on the unit square. This will again be slightly more involved.

3.5 Convergence speed of MLEs

Getting to the final conclusion that the error of the MLE is $O_p(1/n)$ will be the same as in one dimension.

The problem in this section will be to prove that the difference between the MLE and the QMLE is $o_p(1/n)$. The idea of the proof will be the same as in one dimension. We compare the log-likelihood evaluated in $(\hat{\lambda}, \hat{\mu}, \hat{\theta})$ with the log-likelihood evaluated in $(\tilde{\lambda}, \tilde{\mu}, \hat{\theta})$. The proof of the theorem will not give many problems. The proof of the lemma where we make use of the order statistics will be a bit more involved than its one-dimensional counterpart, Lemma 2.19. The observations can not be ordered directly, as they do not lie on a straight line. However, it will be possible to order them based on their (horizontal) distance to the line induced by (λ_0, μ_0). In one dimension, we had to flip the two parts of the unit interval and stretch them in order to create a uniform distribution where the points closest to the boundary were on the outside. Something similar to this can likely be done. However, this time the stretching might need to be done in multiple directions. After pasting the backsides of the two areas with different density together, there is a vertical boundary in the middle. The two trapezoids need to be stretched in a way to make sure that the horizontal perturbation from the left boundary is uniformly distributed.

Chapter 4

Concluding remarks

In the introduction it was hypothesised that the MLE for the boundary in the two-dimensional model is $O_p(1/n)$ as $n \to \infty$. This claim has not been proven, but it does seem very likely. What has been proven is that the MLE for the boundary in the one-dimensional model is $O_p(1/n)$ as $n \to \infty$. The main goal of working out this proof was to get a better idea of the intricacies that are involved within this proof. This would then allow us to generalise the proof to two dimensions. The proof of the one-dimensional problem turned out to contain many more details than initially expected. This makes it even more helpful for tackling the two-dimensional version.

There is even a falsely stated lemma in the paper by Chernoff and Rubin[3]. Lemma 4 from the paper, which corresponds to Lemma 2.17, is wrong in saying that δ does not need to depend on ϵ. This statement is stronger than necessary and it does not hold. It seems as if it was stated in the paper, because it shortens the lemma statement. These types of mistakes are harder to spot if the proof is not worked out before trying to generalise it.

The proof of the one-dimensional case helped create a promising start to the proof of the two-dimensional case. However, there are still many gaps that have to be filled in order to prove the hypothesis.

Appendix A

Omitted proofs

This appendix includes proofs that have been omitted in the text. These proofs were moved to the appendix to keep the focus of the main text on the proofs that are most relevant to the two-dimensional generalisation. The first two sections will give proofs from Chapter 2. The last section gives a proof from Chapter 3.

A.1 Proof of Lemma 2.5

Lemma 2.5 states that if G_n is the eCDF for a random variable which is uniformly distributed on $[0, 1]$, then

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = O_p(1), \tag{A.1.1}$$

for $n \to \infty$.

Before we give the complete proof, we give a heuristic overview. First we use the reverse triangle to see that

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| \leq \sup_{x\in(0,1)} \left|\frac{G_n(x)}{x} - 1\right| + 1. \tag{A.1.2}$$

Hence it suffices to show that

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x} - 1\right| = O_p(1). \tag{A.1.3}$$

We bound this in two parts. First we bound $\left|\frac{G_n(x)}{x} - 1\right|$ on the interval $[0, a/n]$ for some conveniently chosen a. To do this we use the probability of G_n being equal to zero at a/n. This probability converges to $\exp(-a)$, so we can simply choose some suitable a. Now we only need to bound G_n on $[a/n, 1]$. Chebyshev's inequality tell us that for any $M > 1$, $n \in \mathbb{N}$, and $x > 0$ we have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \geq M\right) \leq \frac{1}{nxM^2}. \tag{A.1.4}$$

We use this to bound $\left|\frac{G_n(x)}{x} - 1\right|$ at the sequence $2^i \cdot a/n$ for $i \in \mathbb{N}$. It tells us that the probability of being larger than M in all points in the sequence is less than $\frac{2}{M^2 a}$. Thus M can be chosen to make this probability arbitrarily small. Finally we can use the following lemma to bound everything between the points.

Lemma A.1. *Suppose that* $x \in (0, 1)$ *such that*

$$\left|\frac{G_n(x)}{x} - 1\right| < M, \quad \text{and} \quad \left|\frac{G_n(2x)}{2x} - 1\right| < M. \tag{A.1.5}$$

Then for any $y \in (x, 2x)$*, we have*

$$\left|\frac{G_n(y)}{y} - 1\right| < 2M + 3. \tag{A.1.6}$$

This makes sure that everything between the points in the sequence is also bounded. With this, we can finally get to the final conclusion. Now we will give a proof with all the details included. After that we will give the proof of Lemma A.1.

Proof of Lemma 2.5. The first step is to bound the supremum from above by some more convenient supremum. We will then show that this new expression is $O_p(1)$ for $n \to \infty$.

Using the reverse triangle inequality, we see that

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = \left|\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right|\right| \tag{A.1.7}$$

$$= \left|\sup_{x\in(0,1)} \left(\left|\frac{G_n(x)}{x}\right| - 1\right) + 1\right| \tag{A.1.8}$$

$$\leq \left|\sup_{x\in(0,1)} \left(\left|\frac{G_n(x)}{x}\right| - 1\right)\right| + 1 \tag{A.1.9}$$

$$\leq \sup_{x\in(0,1)} \left|\frac{G_n(x)}{x} - 1\right| + 1. \tag{A.1.10}$$

Hence it would suffice to show that

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x} - 1\right| = O_p(1). \tag{A.1.11}$$

Let $\epsilon > 0$. We want to find $M > 0$ and $N \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x} - 1\right| > M\right) < \epsilon. \tag{A.1.12}$$

We will take a couple of steps in order to prove the result. The steps are roughly as follows: First we stochastically bound $\left|\frac{G_n(x)}{x} - 1\right|$ by 0 on $[0, a/n]$ for some conveniently chosen $a \in \mathbb{R}$. Then we will bound it on a sequence of points starting at a/n. Finally, we will show that this also bounds everything between the points.

Define

$$a := -\log\left(1 - \frac{\epsilon}{4}\right). \tag{A.1.13}$$

Let $N_1 \in \mathbb{N}$ such that for all $n > N_1$ we have $a/n < 1$. Then for $n > N_1$ we have

$$\begin{aligned}
\mathbb{P}\left(G_n\left(\frac{a}{n}\right) = 0\right) &= \mathbb{P}\left(\forall_{i \in \{1,2,\dots,n\}} : X_i > \frac{a}{n}\right). && \text{(A.1.14)}\\
&= \left(1 - \frac{a}{n}\right)^n && \text{(A.1.15)}\\
&\to \exp(-a) && \text{(A.1.16)}\\
&= 1 - \frac{\epsilon}{4}. && \text{(A.1.17)}
\end{aligned}$$

Thus we can choose $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\mathbb{P}\left(G_n\left(\frac{a}{n}\right) = 0\right) > 1 - \frac{\epsilon}{2}. \tag{A.1.18}$$

Since G_n is increasing, we find that

$$\mathbb{P}\left(\sup_{x \in (0, a/n)} G_n(x) = 0\right) > 1 - \frac{\epsilon}{2}. \tag{A.1.19}$$

We now move on to bounding G_n on the rest of the domain. Recall Chebyshev's inequality from Theorem B.5. We want to apply Chebyshev's inequality on $\frac{G_n(x)}{x}$. To do this, we first compute the expectation of $G_n(x)/x$. For any $x \in (0, 1)$ we find that

$$\begin{aligned}
\mathbb{E}\left[\frac{G_n(x)}{x}\right] &= \frac{1}{x}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_i \leq x}\right] && \text{(A.1.20)}\\
&= \frac{1}{nx}\sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}_{Y_i \leq x}\right] && \text{(A.1.21)}\\
&= \frac{1}{nx}\sum_{i=1}^{n} \mathbb{P}\left(Y_i \leq x\right) && \text{(A.1.22)}\\
&= 1. && \text{(A.1.23)}
\end{aligned}$$

By Chebyshev's inequality, we know that for all $M > 0$ and $x \in (0, 1)$ we have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \geq M\right) \leq \frac{\sigma^2}{M^2}, \tag{A.1.24}$$

where σ^2 is the variance of $G_n(x)/x$. We compute that for any $n \in \mathbb{N}$ and $x \in (0, 1)$ the variance is given by

$$\sigma^2 = \mathbb{E}\left[\left(\frac{G_n(x)}{x}\right)^2\right] - \mathbb{E}\left[\frac{G_n(x)}{x}\right]^2 \tag{A.1.25}$$

$$= \frac{1}{x^2n^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}_{Y_i\le x}\mathbf{1}_{Y_j\le x}\right] - 1 \tag{A.1.26}$$

$$= \frac{1}{x^2n^2}\left(\mathbb{E}\left[\sum_{i=1}^{n}\mathbf{1}_{Y_i\le x}\right] + \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}_{Y_i\le x}\mathbf{1}_{Y_j\le x}\right]\right) - 1 \tag{A.1.27}$$

$$= \frac{1}{x^2n^2}\left(nx + (n^2-n)x^2\right) - 1 \tag{A.1.28}$$

$$= \frac{1}{nx} - \frac{1}{n} \tag{A.1.29}$$

$$\le \frac{1}{nx}. \tag{A.1.30}$$

Therefore for all $M > 0$ and $x \in (0,1)$ we have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \ge M\right) \le \frac{1}{nxM^2}, \tag{A.1.31}$$

for all $n \in \mathbb{N}$. For all $x > 1$ and $n \in \mathbb{N}$ we have

$$\frac{G_n(x)}{x} = \frac{1}{x} \in (0,1). \tag{A.1.32}$$

Therefore for $x > 1$ and $n \in \mathbb{N}$ we have

$$\left|\frac{G_n(x)}{x} - 1\right| \le 1. \tag{A.1.33}$$

Hence for $x, M > 1$ and $n \in \mathbb{N}$ we also have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \ge M\right) = 0 \le \frac{1}{nxM^2}. \tag{A.1.34}$$

Hence for any $M > 1$ and $x > 0$ we have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \ge M\right) \le \frac{1}{nxM^2}, \tag{A.1.35}$$

for all $n \in \mathbb{N}$. Now we can bound $\left|\frac{G_n(x)}{x} - 1\right|$ at a sequence of points. For any $n \in \mathbb{N}$ we consider the sequence $2^i a/n$. By the Fréchet inequalities we find that for $M > 1$ and $n \in \mathbb{N}$

we have

$$\mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|\geq M\right)\leq\sum_{i=0}^{\infty}\mathbb{P}\left(\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|\geq M\right) \tag{A.1.36}$$

$$\leq\sum_{i=0}^{\infty}\frac{1}{M^2a2^i} \tag{A.1.37}$$

$$=\frac{2}{M^2a}. \tag{A.1.38}$$

We fix $M>1$ such that for any $n\in\mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|\geq M\right)<\frac{\epsilon}{2}. \tag{A.1.39}$$

We shall now use Lemma A.1 and the other previous results to show that for n large enough

$$\mathbb{P}\left(\sup_{x\in(0,1)}\left|\frac{G_n(x)}{x}-1\right|<2M+3\right)<\epsilon. \tag{A.1.40}$$

Lemma A.1 tells us that for any $n\in\mathbb{N}$ we have

$$\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|<M\implies\sup_{x\in(a/n,1)}\left|\frac{G_n(x)}{x}-1\right|<2M+3. \tag{A.1.41}$$

Hence for $n\in\mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|<M\right)\leq\mathbb{P}\left(\sup_{x\in(a/n,1)}\left|\frac{G_n(x)}{x}-1\right|<2M+3\right), \tag{A.1.42}$$

and thus

$$\mathbb{P}\left(\sup_{x\in(a/n,1)}\left|\frac{G_n(x)}{x}-1\right|\geq2M+3\right)\leq\mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(2^ia/n)}{2^ia/n}-1\right|\geq M\right)\leq\frac{\epsilon}{2}. \tag{A.1.43}$$

By the Fréchet inequalities, we find that for $n>N$

$$\mathbb{P}\left(\sup_{x\in(0,1)}\left|\frac{G_n(x)}{x}-1\right|<2M+3\right)\leq\mathbb{P}\left(\sup_{x\in(0,a/n)}G_n(x)\geq2M+3\right) \tag{A.1.44}$$

$$+\mathbb{P}\left(\sup_{x\in(a/n,1)}\left|\frac{G_n(x)}{x}-1\right|\geq2M+3\right) \tag{A.1.45}$$

$$\leq\epsilon. \tag{A.1.46}$$

Therefore we can conclude that

$$\sup_{x\in(0,\alpha_0)} \left|\frac{G_n(x)}{x} - 1\right| = O_p(1), \tag{A.1.47}$$

for $n \to \infty$. Hence we can conclude that indeed

$$\sup_{x\in(0,1)} \left|\frac{G_n(x)}{x}\right| = O_p(1), \tag{A.1.48}$$

for $n \to \infty$. ■

In the proof of Lemma 2.5 we made use of Lemma A.1. The proof of that lemma is given below and concludes this section.

Proof of Lemma A.1. Choose x according to the hypothesis and let $y \in (x, 2x)$. Then

$$\left|\frac{G_n(y)}{y} - 1\right| \leq \frac{G_n(y)}{y} + 1. \tag{A.1.49}$$

Since $G_n(x)$ is monotonically increasing and $1/x$ is decreasing, we have

$$\frac{G_n(y)}{y} + 1 \leq \frac{G_n(2x)}{x} + 1 = 2 \cdot \frac{G_n(2x)}{2x} + 1. \tag{A.1.50}$$

However,

$$\frac{G_n(2x)}{2x} < M + 1, \tag{A.1.51}$$

so we conclude that

$$\left|\frac{G_n(y)}{y} - 1\right| < 2M + 3. \tag{A.1.52}$$

■

A.2 Proof of Lemma 2.12

Recall that Lemma 2.12 says that for all $\epsilon, k > 0$ there exist $K > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\mathbb{P}\left(\sup_{x>K/n} \left|\frac{G_n(x)}{x} - 1\right| < k\right) > 1 - \epsilon, \tag{A.2.1}$$

where G_n is the eCDF for the uniform distribution on $[0, 1]$.

This proof is similar to the proof of Equation (A.1.3) from the proof of Lemma 2.5. We start by a short discussion of the difference between the results. The first key difference is that this time we can not choose the bound within the probability, which we can do when proving that something is O_p. For this reason, no part of the proof of that lemma shows the result from this lemma. The second key difference is that this time we can choose K based on k. In the proof of Lemma 2.5, we split the interval $(0, 1)$ into two parts: $(0, a/n]$ and $(a/n, 1)$. The proof involving the latter is what is similar to the proof of this lemma. However, this a was chosen in such a way that the first interval was stochastically bounded and could not be changed in the proof of the second interval. Therefore the proof of this lemma does not show the result necessary in the proof of Lemma 2.5.

The proof strategy will again include stochastically bounding a sequence of points and showing that this implies that everything between the points is bounded too. The difference with the proof of lemma 2.5 will be that we have to choose the distance between the points of the sequence based on k. We will now give a short overview of the proof. Afterwards we give the detailed proof.

By Chebyshev's inequality we know that for any $x, k_0 > 0$ and $n \in \mathbb{N}$

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \geq k_0\right) \leq \frac{1}{nxk_0^2}. \tag{A.2.2}$$

We will bound the supremum of $\left|\frac{G_n(x)}{x} - 1\right|$ by first bounding it on the sequence $(b^i K/n)_{i\in\mathbb{N}}$, where $b > 1$ and $K > 0$. The probability of being larger than some $k_0 > 0$ in all of these points is less than $1/(Kk_0^2(1 - 1/b))$. We can choose b and k_0 such that it is also bounded between the points. Then we choose K such that the probability is small enough, completing the proof.

Proof of Lemma 2.12. Fix $\epsilon, k > 0$. We start by bounding a sequence of points. In the proof of Lemma 2.5, we used Chebyshev's inequality to show Equation (A.1.31). This result tells us that for any $x, k_0 > 0$ and $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\left|\frac{G_n(x)}{x} - 1\right| \geq k_0\right) \leq \frac{1}{nxk_0^2}. \tag{A.2.3}$$

Let $b > 1$ and $K, k_0 > 0$. Then for any $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(b^iK/n)}{b^iK/n} - 1\right| \geq k_0\right) \leq \sum_{i=0}^{\infty}\mathbb{P}\left(\left|\frac{G_n(b^iK/n)}{b^iK/n} - 1\right| \geq k_0\right) \tag{A.2.4}$$

$$\leq \sum_{i=0}^{\infty}\frac{1}{b^iKk_0^2} \tag{A.2.5}$$

$$= \frac{1}{Kk_0^2(1 - 1/b)}. \tag{A.2.6}$$

Now we want to bound everything between the points as well. If for some $x > 0$ and $n \in \mathbb{N}$ we have

$$\left|\frac{G_n(x)}{x} - 1\right| < k_0, \quad \text{and} \quad \left|\frac{G_n(bx)}{bx} - 1\right| < k_0, \tag{A.2.7}$$

then for $y \in [x, bx]$ we have

$$\frac{G_n(y)}{y} - 1 \leq \frac{G_n(by)}{y} - 1 \tag{A.2.8}$$

$$= b\left(\frac{G_n(by)}{by} - 1\right) + b - 1 \tag{A.2.9}$$

$$< bk_0 + b - 1, \tag{A.2.10}$$

and

$$\frac{G_n(y)}{y} - 1 \geq \frac{G_n(y)}{by} - 1 \tag{A.2.11}$$

$$= \frac{1}{b}\left(\frac{G_n(y)}{y} - 1\right) + \frac{1}{b} - 1 \tag{A.2.12}$$

$$> -\frac{k_0}{b} + \frac{1}{b} - 1. \tag{A.2.13}$$

We want to have

$$\left|\frac{G_n(y)}{y} - 1\right| < k, \tag{A.2.14}$$

for all $y \in [x, bx]$ Thus we want to fix $b > 1$ and $k_0 > 0$ such that

$$bk_0 + b - 1 < k, \tag{A.2.15}$$

and

$$-\frac{k_0}{b} + \frac{1}{b} - 1 > -k. \tag{A.2.16}$$

We could for instance fix $b = 1 + k/2$. Then

$$bk_0 + b - 1 = (k_0 + 1)(1 + k/2) - 1 = \frac{k_0 k}{2} + a + \frac{k}{2}, \tag{A.2.17}$$

which is less than k if and only if

$$k_0 < \frac{k}{k+2}. \tag{A.2.18}$$

We also have

$$-\frac{k_0}{b} + \frac{1}{b} - 1 = \frac{1 - k_0}{1 + k/2} - 1, \tag{A.2.19}$$

which is greater than $-k$ if and only if

$$k_0 < \frac{k}{2}(k+1). \tag{A.2.20}$$

Since

$$\frac{k}{2}(k+1) > 0, \quad \text{and} \quad \frac{k}{k+2} > 0, \tag{A.2.21}$$

we can fix

$$k_0 \in \left(0, \min\left\{\frac{k}{2}(k+1), \frac{k}{k+2}\right\}\right). \tag{A.2.22}$$

With the given definitions of b and k_0, we now find that if for some $x > 0$ and $n \in \mathbb{N}$ Equation (A.2.7) holds, then for any $y \in [x, bx]$ we have

$$\left|\frac{G_n(y)}{y} - 1\right| < k. \tag{A.2.23}$$

All that is left is to fix K such that we have the right bound on the probability.

Fix

$$K > \frac{1}{\epsilon k_0^2(1-1/b)}. \tag{A.2.24}$$

Then

$$K > 0, \quad \text{and} \quad \frac{1}{Kk_0^2(1-1/b)} < \epsilon. \tag{A.2.25}$$

If for some $n \in \mathbb{N}$ we have

$$\sup_{i \in \mathbb{N}\cup\{0\}} \left|\frac{G_n(b^i K/n)}{b^i K/n} - 1\right| < k_0, \tag{A.2.26}$$

then for all $i \in \mathbb{N}$ we have

$$\left|\frac{G_n(b^{i-1}K/n)}{b^{i-1}K/n} - 1\right| < k_0, \quad \text{and} \quad \left|\frac{G_n(b^i K/n)}{b^i K/n} - 1\right| < k_0, \tag{A.2.27}$$

and therefore using Equation (A.2.23) we find that

$$\sup_{x > K/n} \left|\frac{G_n(x)}{x} - 1\right| < k. \tag{A.2.28}$$

Thus for any $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\sup_{x>K/n}\left|\frac{G_n(x)}{x} - 1\right| < k\right) \geq \mathbb{P}\left(\sup_{i\in\mathbb{N}\cup\{0\}}\left|\frac{G_n(b^iK/n)}{b^iK/n} - 1\right| < k_0\right) \tag{A.2.29}$$

$$\geq 1 - \frac{1}{Kk_0^2(1-1/b)} \tag{A.2.30}$$

$$> 1 - \epsilon. \tag{A.2.31}$$

■

A.3 Maximum likelihood estimators in 2D

In this section we will give the proof of the fact that

$$(\hat{\lambda}_n, \hat{\mu}_n) := \underset{(x_1,x_2)\in(0,1)^2}{\arg\max} \Bigg(\log\left(\frac{F_n(x_1,x_2)}{\alpha(x_1,x_2)}\right) F_n(x_1,x_2) + \log\left(\frac{1-F_n(x_1,x_2)}{1-\alpha(x_1,x_2)}\right)(1-F_n(x_1,x_2))\Bigg), \tag{A.3.1}$$

and

$$\hat{\theta}_n := F_n(\hat{\lambda}_n, \hat{\mu}_n), \tag{A.3.2}$$

are the MLEs for λ_0, μ_0 and θ_0, as was mentioned in Section 3.1. This proof is completely analogous to Section 2.1. To compute the MLEs, we first look at the likelihood

$$\begin{aligned}
\mathcal{L}_n(\lambda,\mu,\theta) &= \prod_{i=1}^n f(X_i,\lambda,\mu,\theta) && \text{(A.3.3)}\\
&= \prod_{X_i\in L(\lambda,\mu)} \beta(\lambda,\mu,\theta) \prod_{X_i\in R(\lambda,\mu)} \gamma(\lambda,\mu,\theta) && \text{(A.3.4)}\\
&= (\beta(\lambda,\mu,\theta))^{nF_n(\lambda,\mu)}(\gamma(\lambda,\mu,\theta))^{n(1-F_n(\lambda,\mu))}. && \text{(A.3.5)}
\end{aligned}$$

Therefore the log-likelihood divided by n is given by

$$\ell_n(\lambda,\mu,\theta) = \log\left(\frac{\theta}{\alpha(\lambda,\mu)}\right) F_n(\lambda,\mu) + \log\left(\frac{1-\theta}{1-\alpha(\lambda,\mu)}\right)(1-F_n(\lambda,\mu)). \tag{A.3.6}$$

First we want to find the MLE for θ. The partial derivative of Equation (A.3.6) with respect to θ is given by

$$\frac{\partial \ell_n}{\partial\theta}(\lambda,\mu,\theta) = \frac{F_n(\lambda,\mu)}{\theta} - \frac{1-F_n(\lambda,\mu)}{1-\theta}. \tag{A.3.7}$$

The zero of this equation is given by $F_n(\lambda,\mu)$. To check whether this is actually the MLE, we compute the second order partial derivative of Equation (A.3.6) with respect to θ and evaluate it in $\theta = F_n(\lambda,\mu)$. This yields

$$\begin{aligned}
\frac{\partial^2 \ell_n}{\partial\theta^2}(\lambda,\mu,F_n(\lambda,\mu)) &= -\frac{1}{F_n(\lambda,\mu)} - \frac{1}{1-F_n(\lambda,\mu)} && \text{(A.3.8)}\\
&= \frac{-1}{F_n(\lambda,\mu)(1-F_n(\lambda,\mu))}. && \text{(A.3.9)}
\end{aligned}$$

We know that $F_n(\lambda,\mu)(1-F_n(\lambda,\mu)) > 0$ for any values of λ and μ. Therefore Equation (A.3.9) is negative. We can conclude that

$$\hat{\theta}_n := F_n(\hat{\lambda}_n, \hat{\mu}_n) \tag{A.3.10}$$

is the MLE for θ_0, where $\hat{\lambda}_n$ and $\hat{\mu}_n$ are the MLEs for λ_0 and μ_0 respectively.

Now we want to find the MLEs for λ_0 and μ_0. To do this, we will fill in the MLE for θ_0 into Equation (A.3.6). This yields

$$\begin{aligned} \ell_n(\lambda, \mu, F_n(\lambda, \mu)) = \log\left(\frac{F_n(\lambda, \mu)}{\alpha(\lambda, \mu)}\right) F_n(\lambda, \mu) \\ + \log\left(\frac{1 - F_n(\lambda, \mu)}{1 - \alpha(\lambda, \mu)}\right) (1 - F_n(\lambda, \mu)). \end{aligned} \tag{A.3.11}$$

Since the MLE maximises this function, we can write

$$\begin{aligned} (\hat{\lambda}_n, \hat{\mu}_n) := \underset{(x_1, x_2) \in (0,1)^2}{\arg\max} \Bigg(\log\left(\frac{F_n(x_1, x_2)}{\alpha(x_1, x_2)}\right) F_n(x_1, x_2) \\ + \log\left(\frac{1 - F_n(x_1, x_2)}{1 - \alpha(x_1, x_2)}\right) (1 - F_n(x_1, x_2)) \Bigg). \end{aligned} \tag{A.3.12}$$

Appendix B

Basic concepts and theorems

Basic analysis and probability theory knowledge is assumed. There are a couple of key concepts and theorems that will be mentioned due to their relevance. First we will look at two ways to characterise limiting behaviour of real functions and give a few properties of these characterisations.

Definition B.1 (Big O)**.** Let $f, g : A \to \mathbb{R}$ where $A \subset \mathbb{R}$. Then we write

$$f(x) = O(g(x)), \tag{B.0.1}$$

for $x \to \infty$ if there exist $M, N > 0$ such that for all $x > N$ we have

$$|f(x)| \leq M \cdot |g(x)|. \tag{B.0.2}$$

When $a \in \mathbb{R}$ we write

$$f(x) = O(g(x)), \tag{B.0.3}$$

for $x \to a$ if there exist $\delta, M > 0$ such that

$$0 < |x - a| < \delta \implies |f(x)| \leq M \cdot |g(x)|. \tag{B.0.4}$$

Definition B.2 (Small O)**.** Let $f, g : A \to \mathbb{R}$ where $A \subset \mathbb{R}$. Then we write

$$f(x) = o(g(x)), \tag{B.0.5}$$

for $x \to \infty$ if for every $\epsilon > 0$ there exists $N > 0$ such that for all $x > N$ we have

$$|f(x)| \leq \epsilon \cdot |g(x)|. \tag{B.0.6}$$

When $a \in \mathbb{R}$ we write

$$f(x) = o(g(x)), \tag{B.0.7}$$

for $x \to a$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \implies |f(x)| \leq \epsilon \cdot |g(x)|. \tag{B.0.8}$$

Remark. Statements with big O or small O notation such as Equation B.0.5 can not truly be interpreted as equalities. It is more appropriate to interpret it as f being an element in the set of all functions that are $o(g)$. This will be important when working with expressions that include such notation, such as the properties in the next theorem.

Theorem B.1 (Some properties of big and small O)**.** *In each of the properties that we will mention, the limit that o and O are refer to will be the same. However, it does not matter what these actually are, as the properties always hold. With f, g, f_1, f_2, g_1, and g_2 we denote functions.*

- *If $g_1 = o(f_1)$ and $g_2 = O(f_2)$, then $g_1 \cdot g_2 = o(f_1 \cdot f_2)$.*
- *If $g = o(f)$, then $g = O(f)$.*
- *If $g_1 = o(f_1)$ and $g_2 = f_2 \cdot g_1$, then $g_2 = o(f_1 \cdot f_2)$. The same property holds for big O.*
- *If $g_1 = o(f)$ and $g_2 = o(f)$, then $g_1 + g_2 = o(f)$. The same property holds for big O.*

We continue by considering random variables. First we will define three relevant types of convergence of sequences of random variables and state a relation between them. Then we consider two ways of characterising asymptotic behaviour of sequences of random variables. These two characterisations are the probabilistic equivalents of big O and small O as defined in Definitions B.1 and B.2. We then proceed by stating some well-known theorems in probability theory that will be useful later on.

Definition B.3 (Convergence in distribution)**.** Let $(X_n)_{n\in\mathbb{N}}$ and X be random variables. Let $(F_{X_n})_{n\in\mathbb{N}}$ and F_X be the corresponding cumulative distribution functions. Then X_n is said to *converge in distribution* to X if for all continuity points $x \in \mathbb{R}$ of F_X we have

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x). \tag{B.0.9}$$

This is also denoted as

$$X_n \xrightarrow{d} X. \tag{B.0.10}$$

Definition B.4 (Convergence in probability)**.** Let $(X_n)_{n\in\mathbb{N}}$ and X be random variables. Then X_n is said to *converge in probability* to X if for all $\epsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0. \tag{B.0.11}$$

This is also denoted as

$$X_n \xrightarrow{p} X. \tag{B.0.12}$$

Definition B.5 (Almost sure convergence)**.** Let $(X_n)_{n\in\mathbb{N}}$ and X be random variables. Then X_n is said to *converge almost surely* to X if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{B.0.13}$$

Theorem B.2. *Almost sure convergence implies convergence in probability and convergence in probability implies convergence in distribution.*

Definition B.6 (Big O in probability)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Then we write

$$X_n = O_p(a_n), \tag{B.0.14}$$

for $n \to \infty$ if for every $\epsilon > 0$ there exist $M > 0$ and $n \in \mathbb{N}$ such that for all $n \geq N$

$$\mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon. \tag{B.0.15}$$

Definition B.7 (Small O in probability)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Then we write

$$X_n = o_p(a_n), \tag{B.0.16}$$

for $n \to \infty$ if for every positive ϵ

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > \epsilon\right) = 0. \tag{B.0.17}$$

Remark. Statements with big O or small O notation such as Equation B.0.16 can not truly be interpreted as equalities. It is more appropriate to interpret it as $(X_n)_{n\in\mathbb{N}}$ being an element in the set of all sequences of random variables that are $o_p(a_n)$. This will be important when working with expressions that include such notation.

Theorem B.3 (Continuous mapping theorem)**.** *Let M and N be metric spaces. Let $(X_n)_{n\in\mathbb{N}}$ and X be random variables defined on M. Suppose that $g : M \to N$ is a function such that $\mathbb{P}(X \in D_g) = 0$ where D_g is the set of discontinuity points of g. Then*

$$X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X), \tag{B.0.18}$$

and

$$X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X). \tag{B.0.19}$$

Definition B.8 (Empirical cumulative distribution function)**.** Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables. Then the *empirical cumulative distribution function* or eCDF for a sample of size $n \in \mathbb{N}$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \leq x}, \tag{B.0.20}$$

and denotes the fraction of the random variables X_i that are less than or equal to any number.

Theorem B.4 (Glivenko-Cantelli theorem)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables sampled from a distribution with CDF $F(x)$. For all $n \in \mathbb{N}$ let $F_n(x)$ be the eCDF for $X_1, \ldots, X_n$. Then*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F| \xrightarrow{a.s.} 0. \tag{B.0.21}$$

Theorem B.5 (Chebyshev's inequality)**.** *Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any $k > 0$*

$$\mathbb{P}\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}. \tag{B.0.22}$$

Theorem B.6 (Fréchet inequalities)**.** *Let $(A_i)_{i\in\mathbb{N}}$ be a sequence of events. Then for all $n \in \mathbb{N}$*

$$\sum_{i=1}^{n} \mathbb{P}(A_i) - (n-1) \leq \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \leq \min\{\mathbb{P}(A_1), \ldots, \mathbb{P}(A_n)\}, \tag{B.0.23}$$

and

$$\max\{\mathbb{P}(A_1), \ldots, \mathbb{P}(A_n)\} \leq \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i). \tag{B.0.24}$$

To finish this section, we will define uniform continuity and state the Heine-Cantor theorem as these might not be known by all students.

Definition B.9 (Uniform continuity)**.** Let (M, d_1) and (N, d_2) be metric spaces. Then a function $f : M \to N$ is called *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in M$ with $d_1(x, y) < \delta$ we have $d_2(f(x), f(y)) < \epsilon$.

Theorem B.7 (Heine-Cantor theorem)**.** *Let M and N be metric spaces. If $f : M \to N$ is continuous and M is compact, then f is uniformly continuous.*

Bibliography

[1] A. Aloi et al. "iPAINT: a general approach tailored to image the topology of interfaces with nanometer resolution". In: *Nanoscale* 8.16 (2016), pp. 8712–8716. DOI: 10.1039/c6nr00445h.

[2] D. van der Haven. "Towards parameterless detection of sub-micron particles and soft interfaces in super-resolution microscopy; an iPAINT analysis method". MA thesis. Eindhoven University of Technology, June 2020.

[3] Herman Chernoff and Herman Rubin. "The Estimation Of The Location Of A Discontinuity In Density". In: *Contribution to the Theory of Statistics* (1956), pp. 19–38. DOI: 10.1525/9780520313880-006.

[4] Herman Chernoff. "Large-Sample Theory: Parametric Case". In: *The Annals of Mathematical Statistics* 27.1 (Mar. 1956), pp. 1–22. DOI: 10.1214/aoms/1177728347.

[5] H. B. Mann and A. Wald. "On Stochastic Limit and Order Relationships". In: *The Annals of Mathematical Statistics* 14.3 (Sept. 1943), pp. 217–226. DOI: 10.1214/aoms/1177731415.

[6] Girish Aras, S.rao Jammalamadaka, and X. Zhou. "Limit distribution of spacings statistics when the sample size is random". In: *Statistics & Probability Letters* 8.5 (1989), pp. 451–456. DOI: 10.1016/0167-7152(89)90026-6.

[7] R. Pyke. "Spacings". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 27.3 (1965), pp. 395–436. DOI: 10.1111/j.2517-6161.1965.tb00602.x.