Universiteit Utrecht

Master of Science Thesis

Estimation of the Period of the Intensity Function of a Cyclical Nonhomogeneous Poisson Process

Author:
Paulo Jorge de Andrade Serra

Supervisor:
Eduard Belitser

November 2008

CONTENTS

1. INTRODUCTION

In this text we will be treating the problem of finding an estimator for the unknown period τ of a periodical intensity function λ of a cyclical Non Homogeneous Poisson Process (NHPP) given a trajectory of one such process.

The main uses for estimating the period (and the intensity function) in the context of NHPP is of course to know the longest amount of time that a phenomenon can go without starting to repeat itself or simply to know how often a phenomenon presents certain types of outcomes. This includes of course all phenomena that can be modelled with NHPP's which occur often in, for example, the fields of communications, hydrology, meteorology, insurance, reliability and seismology[14].

Assuming we are indeed in a cyclical setting, the estimation of the period of the intensity function is closely related with the estimation of the intensity function itself: fully knowing the intensity function will allow us to know the period while the period or an estimator for the period is usually assumed to be know when estimating the intensity function [14]; this interconnection tells us that the problem of estimating τ is of no less importance than that of estimating λ itself.

Knowing the period (if one exists) of a deterministic phenomenon is an important step in studying it. It gives the minimal length that a time window needs to have to capture the full extent of the phenomenon's behaviour. When a step is taken towards stochastic processes used to model phenomena, we typically only have access to trajectories of the process within a certain time window. There, our knowledge of the period will be quite useful since it will allow us, in the NHPP case, to partition the trajectory into smaller trajectories sharing a common distribution – this is quite important, crucial even, in estimating the intensity function. When the period is unknown it is still possible to estimate the intensity function but an estimator for the period is required, and there are requirements on the speed of convergence of one such estimator[14]. The interdependence between estimation of the period and estimation of the intensity function is of course two-way: specific knowledge about the shape of the intensity function will be useful in estimating the period since even though we only interested in how often the phenomenon returns to a behaviour already presented knowing what that behaviour is helps. The problems are therefore quite interconnected and a cooperative take on them seems to be the best: a good estimate for the period will help obtain a better estimate for the intensity function and a better estimate for the intensity function will give useful information for estimating the period.

While estimators are know for this quantity [11, 15, 17] the authors do not insert the estimators into any know specific context making the derivation of its properties rather elaborate; they either do not present experimental data on its performance or simply derive no properties. We try here to make up for this by developing an estimator tha closely follows the principles behind M-estimation and presenting some properties and experimental results for it.

The text will be divided as follows.

We begin by presenting in Chapter 2 a summary of existing estimators for this quantity.

We proceed by presenting the basic definitions that will be used throughout this text followed by some results necessary to establish properties for our estimator – this is the content of the next section of the text: Chapter 3.

We will be proposing two versions of the estimator. In a first approach we will assume we have a sample independent NHPP in a fixed time interval sharing a common intensity function

– this will allow us to derive some properties. In a second approach we assume only to have one observation of the NHPP over a time window with length tending to infinity. For this case we will be presenting an estimator that we will obtain from criterion functions obtained when splitting the trajectory of the process into intervals of a fixed length; we thus decompose the original NHPP into several processes of the same type but with different intensity functions (each a shifted version of the intensity function of the original process). By doing this we will loose the assumption of a common distribution that we had in the first approach. This will be the content of Chapter 4 and Chapter 5 respectively. [CHANGE IF THEY ARE MERGED]

We will also present (in Chapter 6) some notes about the implementation of the algorithm and experimental results that were obtained from such implementation using it with generated data. We also experiment with the some parameters and quantities that might influence the quality of the estimates to get an idea of the quality of the estimator and compare these results as well with those available in the literature.

We close by presenting some conclusions in Chapter 7.

2. ESTIMATION OF THE PERIOD OF A NHPP: A BRIEF SURVEY

Literature specifically about the estimation of the period of a cyclic NHPP is not abundant. Most references to the problem appear in the context of the estimation of the intensity function of these processes. In this context, either the period itself or and estimator for the period is assumed to be known;while the first assumption will not be verified in the generality of cases, the latter does not seem to be a better alternative since the current estimators are all but ideal

In the current chapter we will be presenting the main ideas of the estimators for the period τ of the intensity function of a NHPP that could be found in the literature. We will preserve the notation presented by the respective authors.

2.1 Periodogram Estimator

As mentioned in [15], Vere-Jones recurred to spectral analysis [1, 10] created in [11] an estimator based on the maximum of the Bartlett periodogram [4]. The estimator was constructed specifically for intensity function of the type

$$\lambda(t) = A \exp\left(\rho \cos(\omega t + \phi)\right) \tag{2.1}$$

having thus a period that can be obtained from ω. The author establishes consistency for the estimator, however as mentioned by Bebbington in [15], if the underlying intensity function differs greatly from (2.1) the results will be less than ideal . This estimator has the obvious advantages and disadvantages of making such an assumption on the intensity function. A parametric approach of this nature to the problem seems a bit unnatural, though: the problem of estimating the intensity function is already dependent on having an estimator for the period; it might not be the best of ideas to do this by making an assumption on the unknown intensity function.

2.2 Non-parametric Estimator

Here Mangku [13, 17] presents a nonparametric estimator for the period τ. Given a single realisation of a NHPP $X(\omega)$ observed in some time window $W_n = [a_n, b_n]$ with length $|W_n| = b_n - a_n$ and a parameter space Θ – an open interval in $\mathbb{R}^+$ with $\tau \in \Theta$ – the following quantity is defined for any $\delta \in \Theta$:

$$Q_n(\delta) := \frac{1}{|W_n|} \sum_{i=1}^{N_{n\delta}} \left(X(U_{\delta,i}) - \frac{1}{N_{n\delta}} \sum_{j=1}^{N_{n\delta}} X(U_{\delta,j}) \right)^2 \tag{2.2}$$

where $N_{n\delta} = \lfloor |W_n|/\delta \rfloor$ the maximum number of adjacent intervals of length δ in the window W_n; the $U_{\delta,i}$ intervals are of the form $[a_n + r + (i-1)\delta, a_n + r + i\delta)$ for some $r \in [0, |W_n| - \delta N_{n\delta}]$ [1]. The estimator is then defined as

$$\hat{\tau}_{n,1} := \arg\min_{\delta \in \Theta} Q_n(\delta) \tag{2.3}$$

[1] the value of r is shown to be of little importance in the estimation procedure.

A more general estimator is then presented; first $k\tau$ is estimated for some integer $k = k_n = o(|W_n|)$ by $k\hat{\tau}_{n,k}$ which is given by

$$k\hat{\tau}_{n,k} := \arg\min_{\delta\in\Theta_k} Q_n(\delta) \tag{2.4}$$

where $\hat{\tau}_{n,k}$ denotes the resulting estimator of τ; $\Theta_k = (\tau_{k,0}, \tau_{k,1})$ is an open interval such that no multiple of τ other than $k\tau$ is in this interval and $\Theta_1 = \Theta$. The author states as well that the identification of one such Θ_k requires previous information about the value of $k\tau$ and that "flat parts" in $Q_n(\delta)$ might cause the estimator $\hat{\tau}_{n,k}$ not to be uniquely determined by (2.4).

The author proceeds to derive the consistency of the estimator for which an interval Θ_k as defined above is assumed to be known at a rate of $|W_n|^{-\gamma}$ with $\gamma < \frac{1}{2}$ and for the modified estimator

$$\hat{\tau}^*_{n,k} := \frac{1}{k}\arg\min_{\delta\in\Theta_k} Q^*_n(\delta), \tag{2.5}$$

where for $\delta \in \Theta_k$

$$Q^*_n(\delta) := Q_n(\delta) + \frac{X(W_n \backslash W_{N_n\delta})}{|W_n|}. \tag{2.6}$$

The estimator (2.5) is also shown to be approximately normal with a variance that decreases as $\int_0^\tau (\lambda(s)-\theta)^2\, ds$ increases, with θ the global intensity which is defined as the limit with $n \to \infty$ of $\frac{\mathbb{E}X(W_n)}{|W_n|}$.

2.3 Heuristic Estimator

Another non-parametric estimator exists. Bebbington and Zitikis discuss in [15] the estimator found by Mangku (Section 2.2) and present a modification for it. Its principle, as presented by the authors, is the following: we consider a realisation of a NHPP X with a τ-periodic intensity function λ; we then take a certain interval or time window $W_n = [a, b_n]$ with a a fixed real number and (b_n) a sequence of real numbers larger than a that tends to infinity as $n \to \infty$; we can, for real r and l fit into this time window W_n K adjacent intervals of length l $I_k = (r+(k-1)l, r+kl]$ (K will of course depend on the length of the time window and the values of r and l). If l takes the value τ (or a multiple thereof) then we expect $X(I_k)$ for $k = 1, \dots, K$ (the number of events within of the intervals) to be approximately the same since $X(I_k)$ can be used to estimate $\mathbb{E}X(I_k) = \int_{I_k} \lambda(x)\, dx$ and these areas should coincide for all k.

The authors then refer the form of the estimator found in [17]:

$$\hat{\tau}_{n,min} := \arg_l \min_{(1)} \min_{(2)} \sum_{k=1}^{K-1} \left(X(I_k) - X(I_{k+1})\right)^2 \tag{2.7}$$

where $\min_{(2)}$ represents the minimum taken over r and $\min_{(1)}$ the minimum over l since the "true" period should be the smallest value that one such l can take while still verifying (2.7).

The authors then present some flaws in (2.7). When looking for intervals I_k for a minimal length l, a large number of these intervals is created (K takes a large value); the differences $X(I_k) - X(I_{k+1})$ while small will however not be null making the sum $\sum_{k=1}^{K-1} \left(X(I_k) - X(I_{k+1})\right)^2$ possibly large near the true period. The authors also state but do not illustrate or prove that the same holds when estimating through the maximisation of the sum of squares errors of the mean $K^{-1}\sum_{k=1}^{K} X(I_k)$.

The authors then proceed to presenting a modified version of (2.7) they believe improves on the idea:

$$\hat{\tau}_{n,max} := 2 \arg_l \max_{(1)} \max_{(2)} \sum_{k=1}^{K-1} \left(X(I_k) - X(I_{k+1})\right)^2 \tag{2.8}$$

The factor 2 is used to compensate the fact that maximum should be obtained at half of the period. The estimator will now favour a length for the intervals such that the difference in number of events is maximal supposedly exploiting the fact that in a cyclical function "higher" and "lower" intensities should alternate.

The authors then try to improve (2.8) based on some experimentation and analysis of the results produced by (2.8): it is referred that the quadratic "error" terms $\left(X(I_k) - X(I_{k+1})\right)^2$ can be replaced by $H\left(X(I_k) - X(I_{k+1})\right)$, for H some non-negative function H; also the number of adjacent intervals considered in the terms of (2.8) can be extended from two to more. It is concluded that the quadratic form of the estimator performs better and that while using three terms in the difference as opposed to two does improve the results of the estimator; further increases do not seem to produce significant improvement.

The authors also present some experimental results for the estimator. It is also mentioned that while the periodogram estimator presented in [11] has good asymptotic properties such as asymptotic normality and is more effective computationally, it proves to perform rather badly when the intensity function contains several peaks – meaning several local maxima. An example is however presented of an intensity function that drives the estimator (2.8) to estimate half of the period.

The main limitations of these last two estimators seems to be that the authors assume that a parameter set exists containing no multiples of the period with the exception of τ but fail to discuss how one should come about identifying this parameter set or even if on *can* be identified. Further, the discussion of points other than multiples of τ that minimise (or maximise, for the second non-parametric estimator) the criterion functions presented by the authors is virtually absent from their works; this point is however quite important, if not crucial in terms of establishing that the parameter being estimated is indeed identifiable.

As for the periodogram estimator, unless we somehow know the shape of the intensity function to be of similar to (2.1), the estimator might simply converge to the wrong value [2] which in essence means we have no way of knowing if the value if correct (as an approximation, naturally).

In Chapter 6 we will be comparing the results presented in [15] for the estimator developed by the authors and also for the one presented in Section 2.1 with those produced by ours.

[2] Check Table 6.7 for an example of this.

3. DEFINITIONS AND RESULTS

In this chapter we will be presenting the definitions and results that serve as a starting point for the definition of our estimator. We note that although some of these results will not be used we will still present them to contextualize this text.

3.1 Definitions

We start by compiling some definitions. We will divide these into definition concerning statistical concepts and stochastic concepts.

3.1.1 Limit Results and Consistency

Definition 3.1 (Convergence in Probability)**.** [3, p.5]
A sequence of random variables $\{X_n, n = 0, 1, \ldots\}$ is said to ***converge in probability*** [1] to a random variable X (eventually degenerate) if for all $\epsilon > 0$ we have

$$\mathbb{P}\left(|X_n - X| > \epsilon\right) \to 0 \tag{3.1}$$

And we write $X_n \xrightarrow{P} X$.

Definition 3.2 (Almost Sure Convergence)**.** [3, p.6]
A sequence of random variables $\{X_n, n = 0, 1, \ldots\}$ is said to ***converge almost surely (a.s.)*** [1] to a random variable X if we have

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1 \tag{3.2}$$

And we write $X_n \to X$ *a.s.*.

Theorem 3.1 (Continuous mapping theorem)**.** *[3, p.7] Let $g : \mathbb{R}^k \mapsto \mathbb{R}^m$ be a mapping which is measurable and continuous at every point of a set C such that $\mathbb{P}(X \in C) = 1$. Then*

1. *If $X_n \rightsquigarrow X$ then $g(X_n) \rightsquigarrow g(X)$*
2. *If $X_n \xrightarrow{P} X$ then $g(X_n) \xrightarrow{P} g(X)$*
3. *If $X_n \to X$ (a.s.) then $g(X_n) \to g(X)$ (a.s.)*

Proof. Check [3, p.8] □

Definition 3.3 (Asymptotic Consistency)**.** [3, p.44]
A sequence of estimators $\{T_n, n = 0, 1, \ldots\}$ is said to be ***asymptotically consistent*** for estimating some θ in a parameter set Θ if

$$T_n \xrightarrow{P} \theta \tag{3.3}$$

[1] We consider here a metric space endowed with the Euclidean metric.

Definition 3.4. [7, p.187]
A function $f(\cdot)$ is said to be $o(g(h))$ – read small order of $g(h)$ – if

$$\lim_{h \to 0} \frac{f(h)}{g(h)} = 0 \tag{3.4}$$

Definition 3.5. [3, p.12]
A sequence of random variables X_n is said to be, for a given sequence of random variables R_n, $o_p(R_n)$ if

$$X_n = Y_n R_n \text{ with } Y_n \xrightarrow{P} 0 \tag{3.5}$$

3.1.2 M- and Z-estimators

Definition 3.6 (Empirical Distribution)**.** [3, p.42]
If P is the marginal law of a random sample $X_1, \ldots, X_n$ which we assume to be identically distributed, then we define the ***empirical distribution*** $\mathbb{P}_n$. We thus have

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \tag{3.6}$$

and

$$\mathbb{P} f = \int f(X) dP \tag{3.7}$$

Definition 3.7 (M-estimator)**.** [3, pp.41–42]
Let us suppose we have a parameter $\theta \in \Theta$ attached to the distribution of observations $\{X_1, X_2, \ldots, X_n\}$. An estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ that is a minimum of set of expressions

$$\Psi_n(\theta) := \mathbb{P}_n \psi_\theta = \frac{1}{n} \sum_{i=1}^{n} \psi_\theta(X_i) \tag{3.8}$$

is called a ***M-estimator***.
We also define

$$\Psi(\theta) := \mathbb{P} \psi_\theta = \mathbb{E}\left[\psi_\theta(X)\right] \tag{3.9}$$

We call ψ_θ a score function, which is known and should be measurable; depending on the context we might refer to Ψ_n as the criterion functions and Ψ the limit of the criterion functions or alternatively we will also refer to Ψ_n as estimates for the criterion function and Ψ the criterion function – whenever there might be doubt we will refer to the function specifically.

Definition 3.8 (Z-estimator)**.** [3, pp.41–42]
Let us suppose we have a parameter $\theta \in \Theta$ attached to the distribution of observations $\{X_1, X_2, \ldots, X_n\}$. An estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ that is a solution of a system of equations

$$\Psi_n(\theta) := \mathbb{P}_n \psi_\theta = \frac{1}{n} \sum_{i=1}^{n} \psi_\theta(X_i) = 0 \tag{3.10}$$

is called a ***Z-estimator*** even though it is sometimes also referred to as an M-estimator in the literature.
We also define

$$\Psi(\theta) := \mathbb{P} \psi_\theta = \mathbb{E}\left[\psi_\theta(X)\right] \tag{3.11}$$

We call ψ_θ a score function, which is known and should be measurable; depending of the context we might refer to Ψ_n as the criterion functions and Ψ the limit of the criterion functions or alternatively we will also refer to Ψ_n as estimates for the criterion function and Ψ the criterion function – whenever there might be doubt we will refer to the function specifically.

3.1.3 Distributions

Definition 3.9 (Poisson Distribution)**.** [5, p.66]
A random variable N taking values in $\mathbb{N}_0$ is said to have a ***Poisson distribution with parameter*** $\boldsymbol{\lambda > 0}$ if

$$\mathbb{P}\left(N = n\right) = e^{-\lambda}\frac{\lambda^n}{n!},\ n = 0, 1, \ldots \tag{3.12}$$

And we write $N \sim P(\lambda)$; we will sometimes write $N(\lambda)$ to refer directly to N as a Poisson random variable with parameter λ.

Definition 3.10 (Skellam Distribution)**.** [8]
Let us consider two random variables N_1 and N_2 with Poisson distribution with parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ respectively. A random variable $D = N_1 - N_2$ taking values in $\mathbb{N}$ is said to have a ***Skellam distribution with parameters*** $\boldsymbol{\lambda_1}$ ***and*** $\boldsymbol{\lambda_2}$. We write $D \sim S(\lambda_1, \lambda_2)$.
We will only consider the case where N_1 and N_2 are independent.

Theorem 3.2. *[8]*
Let $D \sim S(\lambda_1, \lambda_2)$, then

$$\mathbb{P}\left(D = n\right) = e^{-(\lambda_1+\lambda_2)}\left(\frac{\lambda_1}{\lambda_2}\right)^{n/2} I_n\left(2\sqrt{\lambda_1\lambda_2}\right),\ n \in \mathbb{N} \tag{3.13}$$

for $I_n(z)$ the modified Bessel function of the first kind:

$$I_n(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos\left(n\theta\right)\, d\theta \tag{3.14}$$

valid for an integer n.

Proof. Proof can be found in [8]. □

Corollary 3.1. *[8]*
Let $D \sim S(\lambda, \lambda)$, then

$$\mathbb{P}\left(D = n\right) = e^{-2\lambda}I_n\left(2\lambda\right),\ n \in \mathbb{N} \tag{3.15}$$

Proof. The equality (3.15) follows directly from (3.13) taking $\lambda_1 = \lambda_2 = \lambda$. □

3.1.4 Processes

Definition 3.11 (Counting Process)**.** [5, pp.288–289]
A stochastic process $\{N(t), t \geq 0\}$ is said to be a ***counting process*** if $N(t)$ represents the total number of events that occur by time t; we thus have

(a) $N(t) \geq 0$ for all $t \geq 0$.

(b) $N(t) \in \mathbb{N}$ for all $t \geq 0$.

(c) If $s < t$ then $N(s) \leq N(t)$.

(d) For $s < t$, $N(t) - N(s)$ represents the number of events that occur in the interval $(s, t]$.

A counting process is said to have ***independent increments*** if the number of events that occur in disjoint time intervals is independent.

A counting process is said to have ***stationary increments*** if the distribution of the number of events that occur in any interval depends only on the length of the interval.

Definition 3.12 (Nonhomogeneous Poisson Process)**.** [5, p.316]

A counting process $\{N(t), t \geq 0\}$ is said to be a ***nonhomogeneous Poisson process (NHPP) with intensity function*** $\lambda(t), t \geq 0$, if

(a) $N(0) = 0$.

(b) The process $N(t)$ has independent increments.

(c) $\mathbb{P}\left(N(t+h) - N(t) \geq 2\right) = o(h)$.

(d) $\mathbb{P}\left(N(t+h) - N(t) = 1\right) = \lambda(t)h + o(h)$.

3.2 Results

In this section we will present the results necessary to build and work with the proposed estimator.

3.2.1 Convergence Results

Theorem 3.3 (Relation Between a.s. Convergence and Convergence in Probability)**.** *[7]*

Let $\{X_n, n = 0, 1, \dots\}$ be a sequence of random variables that converges a.s. to a random variable X; then the same sequence converges in probability to X as well.

Proof. Check [7, p.129]. □

Theorem 3.4 (Strong Law of Large Numbers (SLLN) for IID samples)**.** *[7] Let $\{X_1, X_2, \dots\}$ be a sequence of independent, identically distributed random variables such that for all n we have $\mathbb{E}[|X|] < \infty$ (read X integrable) then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}[X] \ a.s.$$

Proof. Check [7, pp.119–120]. □

Theorem 3.5 (Strong Law of Large Numbers for Independent Sample)**.** *[7] Let $\{X_1, X_2, \dots\}$ be a sequence of independent random variables such that for all n we have $\mathbb{E}[X_n] = 0$ and*

$$\sum_{n=1}^{\infty} \frac{var(X_n)}{n^2}$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = 0 \ a.s.$$

Proof. Check [7, p.118]. □

Theorem 3.6 (Central Limit Theorem (CLT))**.** *[7] Let $\{X_1, X_2, \dots\}$ be an IID sequence of random variables distributed like X where $\mathbb{E}[X] = 0$ and $\sigma^2 := var(X) < \infty$*

Then, if we define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\sqrt{n}\left(\bar{X} - \mathbb{E}[X]\right) \rightsquigarrow N(0, \sigma^2)$$

Proof. Check [7, p.189]. □

Theorem 3.7 (Lyapunov's Central Limit Theorem (Lyapunov's CLT)). *[12] Let $\{X_1, X_2, \dots\}$ be a sequence of independent random variables defined in the same probability space. Let us assume that for all $i \in \mathbb{N}$ both $\mathbb{E}[X_i]$ and $\sigma_i^2 := var(X)$ are defined and finite. Then, if the random variable satisfy the Lyapunov's condition:*

$$\lim_{n\to\infty} \frac{\sqrt[3]{\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}[X_i]|^3}}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = 0$$

we have that

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}[X_i]}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \rightsquigarrow N(0,1)$$

Theorem 3.8 (Lindenberg's Central Limit Theorem (Lindenberg's CLT)[2]). *[12] Let $\{X_1, X_2, \dots\}$ be a sequence of independent random variables defined in the same probability space. Let us assume that for all $i \in \mathbb{N}$ both $\mathbb{E}[X_i]$ and $\sigma_i^2 := var(X)$ are defined and finite. Then, if the random variable satisfy the Lindenberg's condition:*

$$\lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}\left[\frac{(X_i - \mathbb{E}[X_i])^2}{\sum_{i=1}^n \sigma_i^2} \;\middle|\; |X_i - \mathbb{E}[X_i]| > \epsilon \sqrt{\sum_{i=1}^n \sigma_i^2} \right] = 0$$

we have that

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}[X_i]}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \rightsquigarrow N(0,1)$$

3.2.2 Consistency of M- and Z-estimators

Theorem 3.9 (Consistency of M-estimators). *[3] Let us consider $\theta \in \Theta$. Let Ψ_n be a random function and Ψ a fixed function of θ such that for every $\epsilon > 0$,*

$$\sup_{\theta\in\Theta} |\Psi_n(\theta) - \Psi(\theta)| \xrightarrow{P} 0 \tag{3.16}$$

$$\inf_{\theta: d(\theta,\theta_0)\geq\epsilon} |\Psi(\theta)| > 0 = \Psi(\theta_0) \tag{3.17}$$

Then, any sequence of estimators $\hat{\theta}_n$ that nearly minimises Ψ_n converges in probability to θ_0.

Proof. Check [3, pp.45–46]. □

Theorem 3.10 (Consistency of Z-estimators). *[3] Let us consider $\theta \in \Theta$. Let Ψ_n be a random function and Ψ a fixed function of θ such that for every $\epsilon > 0$,*

$$\sup_{\theta\in\Theta} |\Psi_n(\theta) - \Psi(\theta)| \xrightarrow{P} 0 \tag{3.18}$$

$$\inf_{\theta: d(\theta,\theta_0)\geq\epsilon} |\Psi(\theta)| \neq 0 = \Psi(\theta_0) \tag{3.19}$$

Then, any sequence of estimators $\hat{\theta}_n$ that are zeros for Ψ_n converges in probability to θ_0.

Proof. Check [3, p.46]. □

[2] We note that the condition for convergence presented in this theorem is weaker than the one presented in Theorem 3.7.

Lemma 3.1 (Condition for Uniform Convergence of Function in Probability)**.** [3] For every θ in a compact metric space Θ let $x \to f_\theta(x)$ be a given measurable function. Suppose that $\theta \to f_\theta(x)$ is continuous for every x and suppose that there exists a function F such that $|f_\theta| \leq F$ and $\mathbb{E}(F) < \infty$. Then

$$\sup_{\theta \in \Theta} |\mathbb{P}_n f_\theta - \mathbb{P} f_\theta| \xrightarrow{P} 0 \tag{3.20}$$

Proof. Check [3, pp.53–54]. □

3.2.3 Normality of M- and Z-estimators

Theorem 3.11 (Asymptotic Normality of M-estimators)**.** *[3] Assume that the map $\theta \mapsto \psi_\theta(x)$ is twice continuously differentiable in a neighbourhood B of θ_0, for every fixed x, with derivatives (with respect to θ) $\dot\psi_\theta(x)$ and $\ddot\psi_\theta(x)$ such that $|\ddot\psi_\theta(x)| \leq f(x)$ for some function f with $\mathbb{P}f < \infty$ for every $\theta \in B$. Furthermore, suppose that $\mathbb{P}\psi^2_{\theta_0} < \infty$, $\mathbb{P}|\dot\psi_{\theta_0}| < \infty$ and $\mathbb{P}\dot\psi_{\theta_0} \neq 0$. If $\hat\theta_n$ are zeros of $\theta \mapsto \Psi_n(\theta)$ that are consistent for a zero θ_0 of $\theta \mapsto \Psi(\theta)$, then the sequence $\sqrt{n}(\hat\theta_n - \theta_0)$ converges in distribution to a normal distribution with mean zero and variance $\mathbb{P}\psi^2_{\theta_0}/(\mathbb{P}\dot\psi_{\theta_0})^2$.*

Proof. Check [3, p.68]. □

Theorem 3.12 (Normality under Lipschitz Condition)**.** *[3] For each θ in an open subset of a Euclidean space, let $x \mapsto \psi_\theta(x)$ be a measurable real valued such that for every θ_1, θ_2 in a neighbourhood of θ_0 and a measurable function $\dot\psi$ such that $\mathbb{P}\dot\psi^2 < \infty$,*

$$||\psi_{\theta_1}(x) - \psi_{\theta_1}(x)|| \leq \dot\psi(x)||\theta_1 - \theta_2||.$$

Assume that $\mathbb{P}||\psi_{\theta_0}||^2 < \infty$ and that the map $\theta \mapsto \mathbb{P}\psi_0$ is differentiable at θ_0, with nonnull derivative v_{θ_0}. If $\mathbb{P}_n\psi_{\hat\theta_0} = o_p(1/\sqrt{n})$, and $\hat\theta_n \xrightarrow{P} \theta_0$, then

$$\sqrt{n}(\hat\theta_n - \theta_0) = -v_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta_0}(X_i) + o_p(1)$$

In particular we have that the sequence $\sqrt{n}(\hat\theta_n - \theta_0)$ is asymptotically normal with mean 0 *and variance $\mathbb{P}\psi^2_{\theta_0}/(v_{\theta_0})^2$*

The theorem is presented here in its unidimensional version.

Proof. Check [3, pp.52–53]. □

Moments

Definition 3.13 (UI)**.** [7, p.127] A sequence X_n of random variables is ***uniformly integrable (UI)*** if for ani $\epsilon > 0$ there exists a finite $K > 0$ such that $\mathbb{E}[|X| \mid |X| > K] < \epsilon, \forall n \in \mathbb{N}$.

We note that this holds if the absolute value of the sequence is dominated by some positive random variable with finite expectation.

Theorem 3.13 (L^1 convergence)**.** *[7, p.131] Let X_n be a sequence of random variables such that for all $n \in \mathbb{N}$ $\mathbb{E}|X_n| < \infty$ and X such that $\mathbb{E}|X| < \infty$. Then $\mathbb{E}|X_n| \to \mathbb{E}|X|$ if and only if the following two conditions are verified: $X_n \xrightarrow{P} X$ and the sequence X_n is UI.*

Proof. Check [7, pp.131–132]. □

Proposition 3.1 (Moments of Poisson Distribution)**.** Let $N \sim P(\lambda)$, $\lambda \leq 0$. Then $\mathbb{E}(N) = \lambda$ and $var(N) = \lambda$.

Proof. Proof can be found in [5, p.66]. □

Theorem 3.14 (Jensen's Inequality). *Let $g : G \to \mathbb{R}$ be a convex function on an open subinterval G of $\mathbb{R}$ and that X is a random variable such that $\mathbb{E}|X| < \infty$, $\mathbb{P}(X \in G) = 1$ and $\mathbb{E}|g(X)| < \infty$. Then, $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$.*

Proof. Proof can be found in [7, p.61]. □

3.2.4 NHPP Results

Theorem 3.15 (Distribution of the Increments of a NHPP). *[5]*

Let $\{N(t), t \geq 0\}$ be a nonhomogeneous Poisson process with intensity function $\lambda(t), t \geq 0$. We have that for all $s, t \in \mathbb{R}_0^+$

$$N(s+t) - N(s) \sim P\left(\int_s^{s+t} \lambda(t)dt\right) \tag{3.21}$$

Proof. Proof can be found in [5, pp.317–318]. □

3.2.5 Other Results

Lemma 3.2 (Integrals of Periodic Functions). Let $T \geq 0$ and $\lambda(t), t \geq 0$ a periodic function with period $\tau \geq 0$. For any $k \in \mathbb{N}$ we have $\forall k \in \mathbb{N} : 0 \leq k\tau \leq T$, $\int_0^{T-k\tau} \lambda(t)dt = \int_{k\tau}^{T} \lambda(t)dt$

Proof. Take $t' = t + k\tau$, $dt'/dt = 1$. We have,

$$\int_0^{T-k\tau} \lambda(t)dt = \int_{k\tau}^{T} \lambda(t' - k\tau)dt' = \int_{k\tau}^{T} \lambda(t')dt'$$

using τ-periodicity of λ. □

Theorem 3.16 (Rearranging Inequality). *[9]*

Let a_n and b_n be two increasing sequences and $\sigma(1), \ldots, \sigma(n)$ any rearrangement of the integers $1, \ldots, n$. We have that

$$a_n b_1 + \cdots + a_1 b_n \leq a_{\sigma(1)} b_1 + \cdots + a_{\sigma(n)} b_n \leq a_1 b_1 + \cdots + a_n b_n$$

Proof. Check [9]. □

4. DEFINITION OF THE ESTIMATOR FOR IID SAMPLE

In this chapter we will be presenting what will in the following be refer to as the *first version* of the estimator. We will construct this estimator by assuming we have an independent sample of NHPP's known on a certain fixed, finite, time interval all sharing a common periodic intensity function. Our estimator will be analogous to a Z-estimator and will be obtained by comparing counts of events within very specific time intervals.

Before we present the expression for the estimator there will be a section (Section 4.1) dedicated to laying out the principle behind the estimator; after this we proceed with the definition of the estimator (Section 4.2); in the remainder of the chapter we will derive some properties of the estimator.

4.1 Principle Behind the Definition of the Estimator

The discussion that we will have in this section extends a bit beyond our present context and has a focus on properties of periodic functions or more specifically, integrals of periodic functions. The reason for this is simple: the data that is available to us to perform the estimation is event times, or equivalently counts for number of events; we know that the mean number of events that occurs within a fixed time interval of a NHPP is an estimator for the expectation of the process within the same time interval [1]; we also have that this same expectation is the integral of the intensity function again over the same time interval. As such, and using the properties of the process as a stepping stone, if we are able to somehow characterise the period in terms integrals of the intensity function we should be able to find an analogue characterisation for the period in terms of number of events (or counts thereof) and as such find an estimator for the period. This is exactly the principle behind the construction of the estimator. We will, in the remainder of this section explicitly state how we will characterise the period in terms of integrals of the intensity function. The discussion is however independent of our present context of NHPP's and so we will simply be looking in this section at λ as any positive τ-periodic function and forget for now that it will be an intensity function.

4.1.1 Integral of Periodic Functions

Let us assume we have some $T > 0$ (that we will first assume, for simplicity, is between τ and 2τ) and that we are interested in looking at λ only in $[0, T]$. If λ repeats itself every τ units then the values that it takes in the intervals $[\tau, T-\tau]$ and $[0, T-\tau]$ should be the same (and trivially also the integrals over these intervals). If we now concentrate ourselves in comparing the integrals of λ over intervals of the form $[0, \theta]$ and $[T-\theta, T]$ (both of the same length) we know that these integrals will coincide at least for $\theta = T - \tau$. In Figure 4.1 we show a scheme for a sinus-type function with period τ.

In the upper left graph in Figure 4.1 we see that the area under the first interval, $[0, \theta]$, is smaller than the area under the second interval, $[T-\theta, T]$, by taking $\theta < T - \tau$; the relation between the areas in inverted in the upper right graph when we take $\theta > T - \tau$; for $\theta = T - \tau$ the bottom left graph shows the repetition of λ in the considered interval; the bottom right graph shows the representation of the difference of the two integrals (integral over first interval minus integral over second interval) as a function of θ for $\theta \in [0, T]$.

[1] The expected number of events within an interval can be trivially estimated by the mean number of events that we observe within that same time interval.

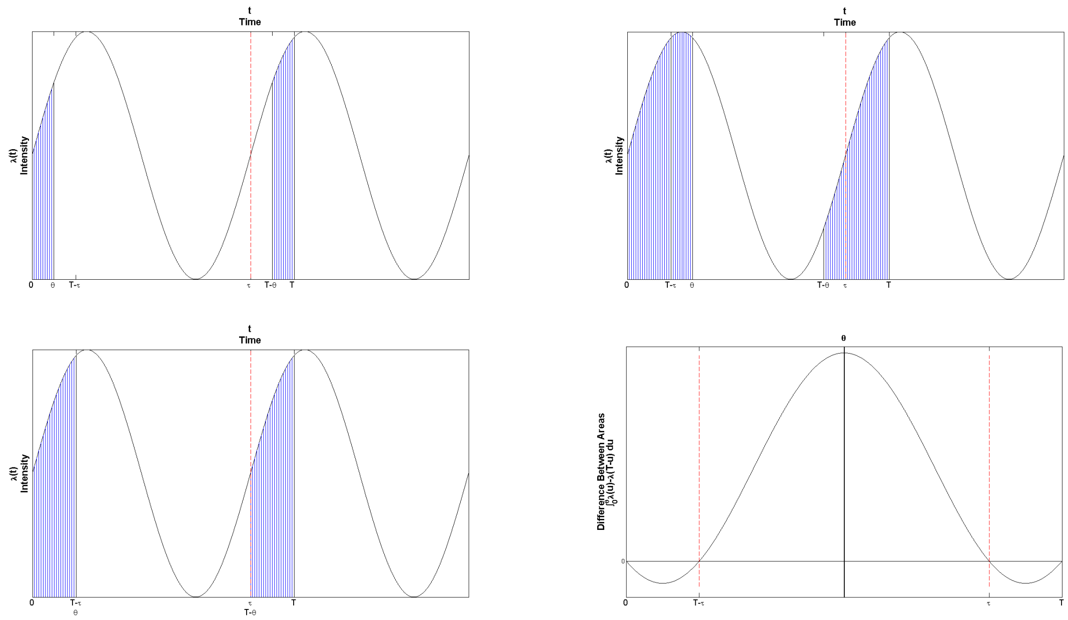


Fig. 4.1: Changes in area under a periodic function over $[0, \theta]$ and $[T - \theta, T]$ for different values θ. The shaded regions are being compared.

Before we discuss with some more detail what we have seen so far let us first introduce some notation[2] for the integrals we are working with:

$$\overrightarrow{n}_T(\theta) := \int_0^\theta \lambda(t)dt \tag{4.1}$$

$$\overleftarrow{n}_T(\theta) := \int_{T-\theta}^T \lambda(t)dt \tag{4.2}$$

The meaning of the arrows in (4.1) and (4.2) – the, say, forward integral and the backward integral – should be evident from looking at Figure 4.1. We will also be referring to the function in the bottom right corner of the same figure as

$$\Delta_T(\theta) := \overrightarrow{n}_T(\theta) - \overleftarrow{n}_T(\theta) \tag{4.3}$$

with all of these three functions being defined for $\theta \in [0, T]$.

As suggested by the situation in Figure 4.1 we note that the (4.3) is symmetrical as a function of θ and that both $\Delta_T(T - \tau)$ and $\Delta_T(\tau)$ are null. These two results are summarised bellow:

Proposition 4.1 (Symmetry of Δ_T)**.** The function Δ_T is symmetrical as a function of θ in $[0, T]$, meaning $\Delta_T(\theta) = \Delta_T(T - \theta)$.

Proof. $\Delta_T(T-\theta) = \overrightarrow{n}_T(T-\theta) - \overleftarrow{n}_T(T-\theta) = \int_0^{T-\theta} \lambda(t)\, dt - \int_\theta^T \lambda(t)\, dt = \int_0^{T-\theta} \lambda(t)\, dt + \int_{T-\theta}^T \lambda(t)\, dt - \int_{T-\theta}^T \lambda(t)\, dt - \int_\theta^T \lambda(t)\, dt = \int_0^T \lambda(t)\, dt - \int_\theta^T \lambda(t)\, dt - \int_{T-\theta}^T \lambda(t)\, dt = \int_0^\theta \lambda(t)\, dt - \int_{T-\theta}^T \lambda(t)\, dt = \overrightarrow{n}_T(\theta) - \overleftarrow{n}_T(\theta) = \Delta_T(\theta)$ □

Proposition 4.2 (Zeros of Δ_T)**.** We have that $\Delta_T(T - \tau)$ and $\Delta_T(\tau)$ are null.

[2] We do not include λ in the notation since in we will always be working with a fixed function λ.

Proof. Due to Proposition 4.1 we only need to prove that one of these two quantities is null. The result follows quite easily for $\Delta_T(\tau)$. We have that $\Delta_T(\tau) = \overrightarrow{n}_T(\tau) - \overleftarrow{n}_T(\tau) = \int_0^\tau \lambda(t)dt - \int_{T-\tau}^T \lambda(t)dt$ but we have that $\int_{T-\tau}^T \lambda(t)dt = \int_0^T \lambda(t)dt - \int_0^{T-\tau} \lambda(t)dt$ which due to τ-periodicity can also be written as $\int_0^T \lambda(t)dt - \int_\tau^T \lambda(t)dt = \int_0^\tau \lambda(t)dt$ making $\Delta_T(\tau) = 0$ $\square$

These results then suggest that it might be possible to reduce the search for the value for the period to finding a zero of a function such as Δ_T in $(0, T/2]$.

The previous discussion concerned a value of T such that $\tau < T < 2\tau$. For $T \leq \tau$ we do not expect Δ_T to have any zeros in $(0, T/2]$, while for $T > 2\tau$ we expect a quantity of zeros corresponding at least to the number of multiples of τ smaller than T (this is an immediate consequence of the fact that if a function is τ-periodic it is also $k\tau$-periodic for any integer k, and by further application of the results seen above). So, if we do not know our periodic function λ (or its period) but do know that T is larger than τ and that Δ_T has only one zero in $(0, T/2]$ we can unequivocally identify the period as $T - \theta_0$ with θ_0 such that $\Delta_T(\theta_0) = 0$.

We may however be in a situation where we cannot know if $T > \tau$ or that we do know that $T > \tau$ but we have multiple zeros for Δ_T. In this case we should still be able to find the period by choosing a "good" value for T after some exploration.

4.1.2 The zeros of Δ_T as a function of T

Our ability of properly identifying the period based on the zeros of Δ_T relies greatly on an adequate choice for the value of T. Choosing a "small" value for T will result in a function Δ_T having no zeros in $(0, T/2]$; a "large" choice for T will result in Δ_T having too many zeros in this same interval. While we know that for each multiple of τ smaller than T we get a zero in Δ_T, there may conceivably exist other points that are also zeros of this function but have no correspondence with a multiple of τ. As such, it might conceivably be impossible to distinguish the zeros these two different types of zeros.

The first step in trying to find the period should therefore be the analysis of the zeros of Δ_T as a function of T. The following notation is introduced:

$$\theta_T^0 := \min \arg_{\theta \in (0, T/2]} \Delta_T(\theta) = 0 \tag{4.4}$$

and

$$\zeta(T) := T - \theta_T^0 \tag{4.5}$$

The minimum in (4.4) is necessary to allow the quantity to be a function in case we have either extra zeros that do not correspond to multiples of τ or simply if $T > 2\tau$. Let us also follow the convention that the minimum of an empty set is infinity to deal with with the cases where we have no zeros in $(0, T/2]$. The value of the period should then be obtained from θ_T^0 if an appropriate choice for T is made; we will see that it is possible to make a good choice by looking at the graph of (4.5).

It is certainly possible to construct for fixed $T = T_0$ examples of τ-periodic functions such that $\Delta_T(\theta)$ is null for a value of θ other than $T - \tau$. It is however better to begin with a "well behaved" periodic function such that the zeros of Δ_T correspond only to multiples of τ. Sinus functions like the one previously presented are such functions. The graph of $\zeta(T)$ for this function is presented in Figure 4.2.

We see the ζ has a staircase shape with steps of length τ and height τ: when T is smaller than τ we have no zeros in Δ_T and as such no value for ζ; for $T \in [\tau, 2\tau]$ we have a unique zero at $T - \tau$ and so θ_T^0 becomes τ for all T in the interval. This is indeed the ideal graph to observe in our circumstances: if we observe for a certain T that $\theta \mapsto \Delta_T(\theta)$ has only one zero (besides θ=0) then that zero is necessarily $T - \tau$ and if this holds for all T such that $\tau \leq T < 2\tau$ then we will indeed observe Figure 4.2. In this case τ can be identified correctly.

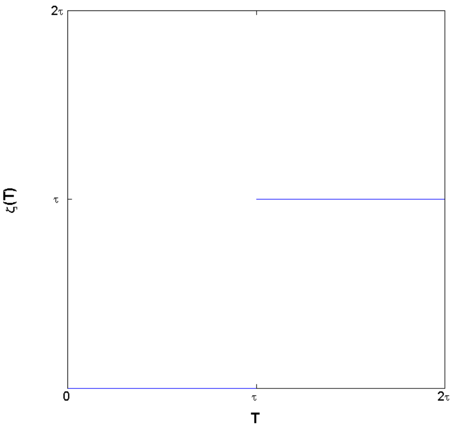


Fig. 4.2: Graph for $\zeta(T)$.

It is also convenient to mention now why in (4.4) the minimum is considered: the first (and most immediate) reason would be so that ζ as defined in (4.5) was a function; a second reason would be that if we look at progressive increases of the value of T, then this value starts to exceed some multiple of τ, say $k\tau$, the fact that we know that there is a zero at $T - k\tau$ means that any other zeros corresponding to other (smaller) multiples of τ will not be considered and so we get a plateau at a level corresponding to the smallest value for θ that is a zero resulting in the highest multiples of τ to be identified.

It is quite simple to cause distortions on the graph of ζ by modifying the τ-periodic function λ. We will be presenting examples of this.

$$\lambda_{T_0}^{\alpha}(t) = \begin{cases} \sin\left(\frac{2\pi}{\tau}(t' - (T_0 - \alpha - \tau))\right) + 1 & \text{if } t' \in [T_0 - \alpha - \tau, T_0 - \tau] \\ \sin\left(\frac{2\pi}{\tau}t'\right) + 1 & \text{if } t' \notin [T_0 - \alpha - \tau, T_0 - \tau] \end{cases} \tag{4.6}$$

for $t' = t - \lfloor \frac{t}{\tau} \rfloor \tau$. We are in essence replacing into the interval $[T_0 - \alpha - \tau, T_0 - \tau]$ the values that the function takes on the interval $[0, \alpha]$; this way, if we take $T' \in (T_0 - \alpha, T_0]$ we have $\Delta_{T'}(\alpha) = 0$. We thus have a family of function for which Δ_T has extra zeros at points other than $T - \tau$. This can be seen, for $\alpha = 1$ and $T_0 = 76$, in Figure 4.3.

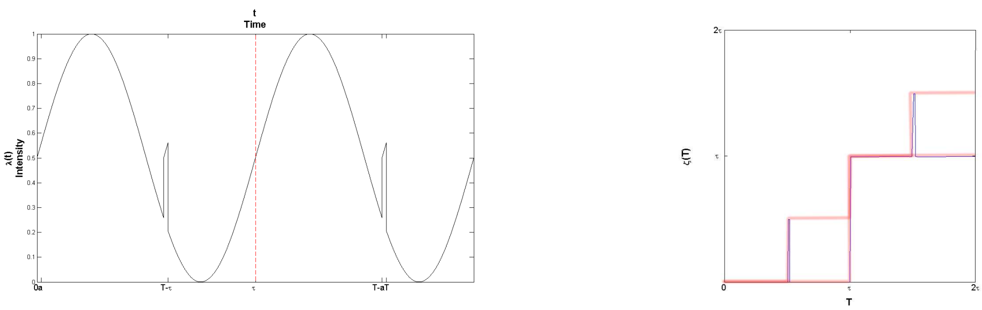


Fig. 4.3: Graph for (4.6) with $\alpha = 1$ and $T_0 = 76$ (left) and the corresponding graph for $\zeta(T)$ (right).

There are several things to notice in Figure 4.3. Firstly, if we concentrate on the interval $[\tau, 2\tau]$ we see that for λ_{76}^{1} we indeed get zeros in Δ_T at points other than $T - \tau$: this explains why we get a smaller step at another height in this interval; $T' - \alpha$ is smaller than $T' - \tau$ for $T' \in (T_0 - \alpha, T_0]$ and so $\zeta(T')$, for T' in the same interval, takes the value $T' - (T' - (T_0 - \alpha)) = T_0 - \alpha$, in our case 75. Secondly, by looking at the interval $[0, \tau]$ we see that the same peak that we saw in $[\tau, 2\tau]$ is now reproduced here due to the periodicity of λ^1_{76}: in essence we are comparing the same values of the function but now the points that provoke the anomalies are shifted back by τ and so $\zeta(T')$, for $T' \in (T_0 - \alpha - \tau, T_0\tau]$, takes the value $T' - (T' - (T_0 - \alpha - \tau)) = T_0 - \tau - \alpha$, in our case 25.

Since these small steps can clearly be seen to be distortions of the larger steps of same length and height, this situation does not pose a problem. Also – and since we know that the function should be constant at values corresponding to the period – the idea that the flat portions of the peaks could indicate actual multiples of the period is unlikely since that would imply those flat portions to be inserted into a staircase-type structure which it is clearly not. Furthermore, we note that these perturbations affect the steps in the function in a periodic fashion further strengthening the identification of the period as not being defined by these "hills". It would in this case be easy to choose a value for T that would help us identify the period correctly: any value in the second large step that is not a part of the small "hill".

We can however see that the identification of an appropriate value for T could be more difficult is we were to further modify λ so that the disturbances would affect larger portions of ζ: ζ would start to look like a staircase function with shorter steps. In Figure 4.4 we can see a trend that could lead to a problematic graph for ζ:

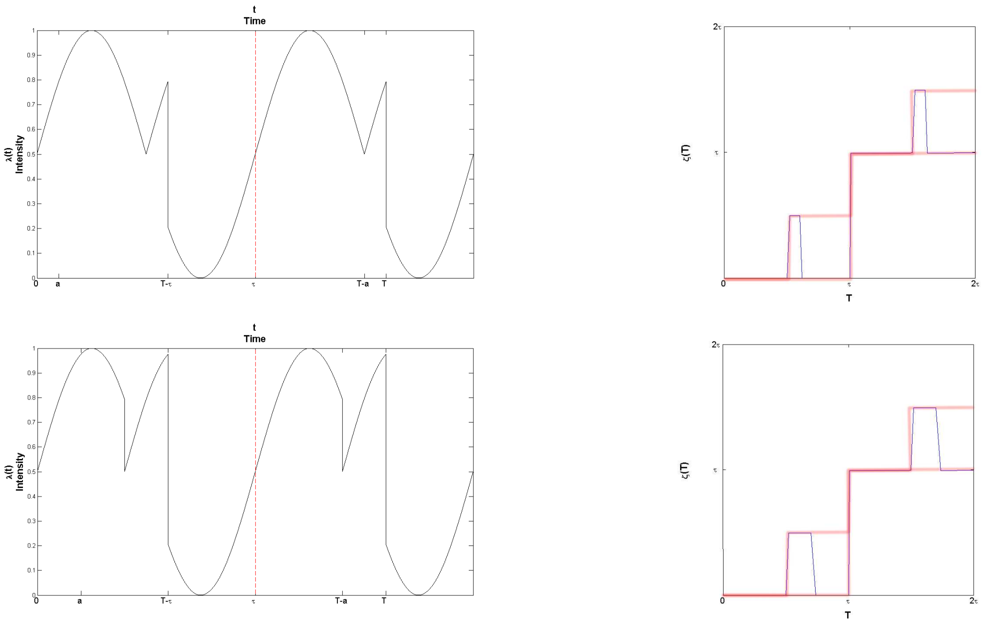


Fig. 4.4: Graph for (4.6) with $\alpha = 5$ and $T_0 = 80$ (upper left) and the corresponding graph for $\zeta(T)$ (upper right) and the same for $\alpha = 10$ and $T_0 = 85$ (bellow).

These modification do seem to be having a toll in the function λ though, since we are repeating the function within its period – note that this sort of repetition is required to maintain the hills as a constant level. It becomes important to see exactly in what way the presence of these extra flat portions actually restrict the shape of λ.

Let us assume we begin with a function ζ such as the one in Figure 4.2 and that we want to introduce two extra steps: one ranging from $\tau/2$ to τ with height $\tau/2$ and another one ranging from $3\tau/2$ to 2τ with height $3\tau/2$, meaning something like the following graph (Figure 4.5).

This would indeed be a way to, while still keeping the steps at 0 and τ introduce extra steps such that we would be driven to draw a wrong conclusion about the value of τ. Note also that this is actually the graph we would expect from a $\tau/2$-periodical intensity function. So, for any value of T_0 in $(3\tau/2, 2\tau)$ we should have that the integral of the function over $[0, T_0 - 3\tau/2]$ and

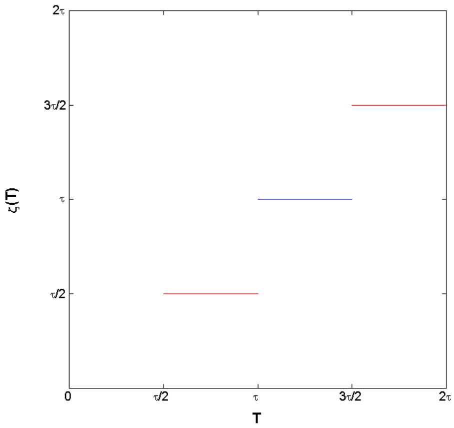


Fig. 4.5: Graph of $\zeta(T)$ after introducing two extra steps (in red).

$[3\tau/2, T_0]$ should coincide this of course implies that λ should take the same values within $[0, \tau/2]$ and $[3\tau/2, 2\tau]$ or, due to the τ-periodicity, within $[0, \tau/2]$ and $[\tau/2, \tau]$. This is certainly possible but it means that the function actually has to be $\tau/2$-periodical; but if we now look back at Figure 4.5, the graph we were trying to obtain is indeed the graph we would expect from a $\tau/2$-periodical function. We thus conclude that requiring there to be an "anomaly" in the graph of ζ such that we might be driven to conclude that the period is actually half of what it really is implies that our function is actually $\tau/2$-periodical, i.e., we would be drawing the right conclusion.

If we were to relax this requirement on the placement and length of these extra steps enough so that this would not imply $\tau/2$-periodicity, we would still have that most of the function would be repeating itself during the two halves of the period; in a case at hand, the graph of ζ would indicate τ-periodicity since the size of the steps (the ones of height 0 and τ and the ones of height $\tau/2$ and $3\tau/2$) would not be the same; also depending on the values that λ takes in the interval(s) where it now does not repeat itself within its period we might even have more steps.

On the other hand, the more we step away from a situation such as the one we just discussed – where we introduce these possibly deceitful steps – the more difficult it is for the extra steps to disguise the underlying τ staircase. We thus see here a sort of *Red Queen effect*[2] where the intensity function seems to fall short of the scrutiny of the function ζ no matter how much we try to choose/modify λ to try and explore the natural shape of ζ.

What was done before can of course be repeated if we try to produce two extra steps per period such as can be seen in Figure 4.6. Utilising exactly the same argument as before we see that the function needs to take the same values in $[0, \tau/3]$ as in $[4\tau/3, 5\tau/3]$ and $[5\tau/3, 2\tau]$ or, by τ-periodicity, in $[0, \tau/3]$, $[\tau/3, 2\tau/3]$ and $[2\tau/3, \tau]$ meaning it has to be $\tau/3$-periodical. The same can of course be extended for any fraction of the type $\tau/m, m \in \mathbb{N}$.

We thus see that the graph of ζ can be safely used as a tool to properly identify good choices for T.

Any extra information that might be known about λ, such as even a very rough estimate for τ or the general shape of λ will of course make the problem of choosing an appropriate T much simpler since using this information, functions mentioned above (namely ζ) can be drawn and the flat areas that characterise safe choices for T can be identified *a prori*. This information, albeit useful, should not however be necessary, as we have seen, in choosing an appropriate value for T.

The next step is to see what was presented so far, namely the integrals (4.2) and (4.1), as the

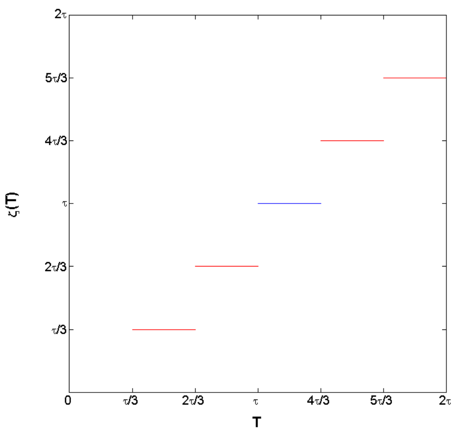


Fig. 4.6: Graph of $\zeta(T)$ after introducing four extra steps (in red).

expected number of events of a NHPP within the considered interval.

4.1.3 Bridging the deterministic case and the NHPP case

Let us assume we have a NHPP $N(t),\ t \geq 0$; we will also assume that the intensity function for this process, $\lambda(t),\ t \geq 0$, is locally integrable and τ-periodic. Let us take some finite $T > 0$ and use a notation

$$\overrightarrow{N}_T(\theta) := N(\theta) \tag{4.7}$$

$$\overleftarrow{N}_T(\theta) := N(T) - N(T-\theta) \tag{4.8}$$

Note that as such, from Theorem 3.15 we have that

$$\overrightarrow{N}_T(\theta) \sim P\left(\int_0^\theta \lambda(t)dt\right) \tag{4.9}$$

$$\overleftarrow{N}_T(\theta) \sim P\left(\int_{T-\theta}^T \lambda(t)dt\right) \tag{4.10}$$

for all $\theta \in [0, T]$. Note also that for $\theta \leq T/2$ the random variables (4.9) and (4.10) are independent by definition of NHPP (Definition 3.12). We further note that

$$\overrightarrow{n}_T(\theta) = \mathbb{E}[\overrightarrow{N}_T(\theta)] \tag{4.11}$$

$$\overleftarrow{n}_T(\theta) = \mathbb{E}[\overleftarrow{N}_T(\theta)] \tag{4.12}$$

If we now define still for $\theta \in [0, T]$

$$D_T(\theta) := \overrightarrow{N}_T(\theta) - \overleftarrow{N}_T(\theta) \tag{4.13}$$

we have that

Proposition 4.3 (Symmetry of D_T). D_T is symmetrical as a function of θ, meaning $D_T(\theta) = D_T(T-\theta)$.

Proof. This is a simple consequence of the definition: $D_T(T-\theta) = \overrightarrow{N}_T(T-\theta) - \overleftarrow{N}_T(T-\theta) = N(T-\theta) - (N(T) - N(T-(T-\theta))) = N(T-\theta) - (N(T) - N(\theta)) = N(\theta) - (N(T) - N(T-\theta)) = \overrightarrow{N}_T(\theta) - \overleftarrow{N}_T(\theta) = D_T(\theta)$ □

Continuing with our parallel with what was seen in the previous sections we also have for $\theta \in [0, T]$

$$\Delta_T(\theta) = \mathbb{E}[D_T(\theta)] \tag{4.14}$$

Since the expectation (4.14), as we have already seen, reduces to a difference of integrals of our intensity function, everything comes full circle and we are in essence in the same situation as we were in Sections 4.1.1 through 4.1.2 as long as we can estimate the mentioned expectation.

It so turns out that this next step of estimating the expectation is more than just a natural step as it places us in the context of z-estimators. In the next sections we will therefore be concerned in explicitly defining a z-estimator for τ.

4.2 Definition of the Estimator

Let us assume we have a family of NHPP $\{N_1, ..., N_n\}$, all sharing a common intensity function $\lambda(t),\ t \geq 0$, on some time interval that contains $[0, T]$ for fixed, finite, $T > 0$. We will consider a sample of stochastic processes $\{D_1, ..., D_n\}$ with each $D_i(\theta) := \overrightarrow{N}_i(\theta) - \overleftarrow{N}_i(\theta)$ using the notation[3] of Section 4.1.3 with θ in some parameter set Θ[4].

Remark 4.1. Since we will be working with a fixed value for T and λ, no indexes are being used to express the dependency of the random variables on these quantities as a way to simplify the notation.

We will now define our estimator by means of the criterion function

$$\Psi_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} D_i(\theta) \tag{4.15}$$

We also have

$$\Psi(\theta) := \mathbb{E}D(\theta) \tag{4.16}$$

Since by (4.9) and (4.10), for all $\theta \in \Theta$ we have that $\mathbb{E}D(\theta) = \int_0^\theta \lambda(t)dt - \int_{T-\theta}^T \lambda(t)dt < \infty$ the strong law of large numbers (Theorem 3.4) assures that (4.16) is well defined (since we are assuming λ integrable) and that for every $\theta \in \Theta$ $\Psi_n(\theta)$ converges almost surely to $\Psi(\theta)$.

We note that although quite similar to a z-estimator, the fact that the parameter θ parametrizes the random variables in our sample and not some score function which is applied to these random variables makes it something different.

It would be our first impulse to define our estimator as follows:

$$\hat{\theta}_n := \arg_{\theta\in\Theta} : \Psi_n(\theta) = 0 \tag{4.17}$$

but the fact that Ψ always has a zero at $T - \tau$ (if $T > \tau$) does not assure that, however, that Ψ_n also has a zero. This might happen for two reasons: Ψ might not change sign in the neighbourhood of $T - \tau$ so our estimates Ψ_n of Ψ might while still converging remain always positive or always negative[5]; even if Ψ changes sign it might still happen that Ψ_n while converging to Ψ (for now pointwise) which is continuous[6] is not continuous itself and so might quite possibly change signs

[3] Since these D_i share a common distribution for all $0 \leq \theta \leq T$ whenever we refer to this common distribution we will drop the index i and say these element have the same distribution as some D.

[4] Even though we are not specifying Θ at this time we must always take it as a subset of $[0, T]$ so that all of the quantities we are working with are well defined.

[5] Check the bottom left graph of Figure 6.25 for an example of an intensity function that causes this for Ψ

[6] This function is the difference of two integrals.

without taking the value zero. In practise we consider a grid of values for θ and θ_n was just taken as a zero of $\Psi_n(\theta)$ in case one exists of the middle point between two consecutive points in the grid of different signs. These two possibilities force us to be more careful in the definition of the estimator: the first version of the estimator will therefore be defined with the help a sequence of estimators $\hat{\theta}_n$ that are near zeros of (4.15):

$$\hat{\theta}_n := \arg_{\theta\in\Theta} : |\Psi_n(\theta)| < a_n \tag{4.18}$$

giving us finally,

$$\hat{\tau} = T - \hat{\theta}_n \tag{4.19}$$

Something must be said about the sequence a_n in (4.18). This sequence must of course converge to zero; how fast it *can* converges to zero is of course dependent on the speed at which Ψ_n converges to Ψ.

The quantities defined so far in this section are quite dependent on our choice of the parameter set Θ. It was already seen in the previous sections of this chapter how our capacity of properly identifying τ relies to great extent on a good choice for T, so (4.19) might not even be properly defined for certain choices for Θ. This will be further discusses in the next sections.

Our situation should still hold a great deal of similarity with z-estimators: (4.15) is defined as a random function that depends on some parameter and that converges to some non random function Ψ, (4.16), that has a zero at a point corresponding to the correct value for the parameter.

4.3 Zeros of the Estimator

Our choice for T and Θ will essentially affect the number of zeros that (4.15) and (4.16) will have. We will, in this section, show that for all $k = 0, 1, \ldots$ such that $T - k\tau \geq 0$, $T - k\tau$ will be a zero of (4.15) and (4.16). We shall prove this in the following lemma (Lemma 4.1):

Lemma 4.1. Let us fix $T > 0$ and consider $k = \in \mathbb{N}$ such that $T - k\tau \geq 0$. We have that for D distributed like any D_i, $\psi_0(D) = 0$, $\Psi_n(T - k\tau) = o_p(1)$ and $\Psi(T - k\tau) = 0$.

Proof. That $\psi_0(D) = 0$ follows from the definition of ψ:

$$\psi_0(D) = D(0) = \overrightarrow{N}(0) - \overleftarrow{N}(0) = N(0) - (N(T) - N(T-0)) = 0$$

Let us now compute[7] $\Psi(T - k\tau)$:

$$\Psi(T-k\tau) = \mathbb{E}D(\theta = T - k\tau) = \mathbb{E}[\overrightarrow{N}(T-k\tau) - \overleftarrow{N}(T-k\tau)] =$$
$$\mathbb{E}[\overrightarrow{N}(T-k\tau)] - \mathbb{E}[\overleftarrow{N}(T-k\tau)] = \int_0^{T-k\tau} \lambda(t)dt - \int_{k\tau}^{T} \lambda(t)dt = 0$$

using (4.9), (4.10) and Lemma 3.2.

Let us now check that $\Psi_n(T - k\tau) = o_p(1)$. We have from above that $\mathbb{E}D(T - k\tau) = 0$ so from (4.15) we see that $\Psi_n(T - k\tau)$ is defined as the mean of n IID random variables with null expectation and therefore, by the strong law of large numbers (Theorem 3.4), converges almost surely to 0 as $n \to \infty$ and is therefore $o_p(1)$. $\square$

We therefore have that the points $T - k\tau$, although unknown, are near solutions the equations $\Psi_n(\theta) = 0$. These are however not the only points with this property. Lemma 4.2 gives us more points with the the same properties as above.

[7] This had already partially been seen in Proposition 4.2.

Lemma 4.2. Let us fix $T > 0$, finite, and consider the time interval $[0, T]$. With D as defined in Section 4.1 we have that $D(t)$ is distributed like $D(T - t)$. We therefore have that under the assumptions of Lemma 4.1, also points of the form $t_0 = k\tau$ and $t_0 = T$ we have $\psi_{t_0}(D) = 0$, $\Psi_n(t_0) = o_p(1)$ and $\Psi(t_0) = 0$.

Proof. This follows immediately from the definition of the process D: $D(t) = N(t) - (N(T) - N(T-t))$, $D(T-t) = N(T-t) - (N(T) - N(t)) = N(t) - (N(T) - N(T-t))$ which is distributed like $D(t)$; and Lemma 4.1. $\square$

Lemma 4.2 has a simple interpretation:

Remark 4.2. the map $t \mapsto \psi_t$ is symmetrical, and therefore so are the maps $t \mapsto \Psi_n(t)$ and $t \mapsto \Psi(t)$.

These last results are not surprising: we had already seen in the first sections of this chapter results about the zeros and symmetries of the functions we were working with that are analogous to these.

There is still another ingredient that is required for all to work properly: we need to justify that the sequence of criterion functions does have a zero (or almost zero) since otherwise neither (4.17) or (4.18) will be defined. We need to make sure that the correct value for the parameter is in the parameter set. This requirement is not specific to our problem, but general in z-estimation; the particularity here is the existence of multiple values that are zeros for the limiting criterion function (including values that might not correspond to multiples of the period): the problem is thus one of identification and not so much one of existence.

So, by the end of this section we have already seen that we are working with a sequence of criterion functions that converges (just pointwise for now) to a non deterministic limiting function that has a zero at the correct value for the parameter (meaning $T - \tau$). We've also seen that at the correct value for the parameter the sequence of criterion functions converges in probability to zero; so, in principle, if we consider a sequence of zeros for the criterion function this sequence should converge to the correct value of the parameter since in the limit it will be the (unique) point where the limiting function is null (the guarantee that this zero is unique can be achieved by a proper choice for the parameter set Θ.) This is the basic argument for consistency whose precise statement and corresponding proof will be presented in the following section.

4.4 On Establishing the Consistency of the Estimator

In this section we will lay the ground to proving the consistency of our estimator. This consistency depends on the value of T and the parameter set Θ: Theorem 3.10 assures that, under certain assumptions, any sequence of estimators that is the unique "almost zero" of the criterion function (4.15) converges in probability to the zero of (4.16). We are however not working with a z-estimator; the conditions required for consistency to hold for a z-estimator should however indicate what conditions are needed in our case: existence and unicity of a zero in the parameter set and a convergence result from $\Psi_n(\theta)$ to $\Psi(\theta)$ that is uniform over θ and a monotonicity of the $\Psi_n(\theta)$ sequence. Under these assumptions the proof for consistency should to great extent mimic its z-estimator counterpart.

For the limiting function Ψ, being continuous in $[0, T/2]$ and under the assumption that we are working on some compact interval Θ contained in $[0, T/2]$ that we would assume contains a unique zero – θ_0 for Ψ – we should be able to reduce Θ to some Θ', a neighbourhood of θ_0 such that Ψ is monotonous in Θ'. We would then use the uniform convergence result to show that, with probability converging to one, the sequence of Ψ_n should also change signs in Θ', an event that is equivalent to the existence of an almost zero, and so the sequence of almost zeros would fall into any neighbourhood of θ_0 with probability converging to one. As for establishing the uniform convergence result, it should follow from the fact that for each n the variance of $\Psi_n(theta)$ increases with θ and as such the supremum over θ of the difference between $\Psi_n(theta)$ and $\Psi(theta)$ would have $|\Psi_n(T/2) - \Psi(T/2)|$ as an upper bound and this quantity still converges in probability to zero.

We also need to think of the case where Θ contains no zeros for Ψ. This can happen under several circumstances: if $T < \tau$ and the intensity function contains no "extra zeros" or even if $T > \tau$ but we failed to include any zeros for Ψ in Θ. In these situation we simply conclude that out parameter set does not contain any zeros for Ψ and we must choose another parameter set.

If we know that there are no extra zeros – say if we have previous knowledge on the shape of λ – there is a very simple way to construct a parameter set as required to establish consistency. We start by trying to estimate the period using for some small $\epsilon > 0$ (say, the smallest inter event time in our sample) and consider any $T = T_0$ and a parameter set $\Theta_1 = [\epsilon, T_0/2]$ (we know this is enough due to the symmetry of the criterion function) if we have zeros in this set we halve T and consider $\Theta_2 = [\epsilon, T_0/2^2]$; we repeat this until we reach $\Theta_m = [\epsilon, T_0/2^m]$ where there are no zeros: our final parameter set should then be Θ_{m-1} with $T = T_0/2^m$. If however for $T = T_0$ and a parameter set $\Theta_1 = [\epsilon, T_0/2]$ we have no zeros we should double T and consider $\Theta_2 = [\epsilon, T_0]$; we repeat this until we reach $\Theta_m = [\epsilon, T_0/2^{m-2}]$ where there is one zero and this should be our parameter set. In both situations if $T_0 < \tau$ then $2T_0 \leq 2\tau$ (and we therefore have at most one zero in $[\epsilon, T_0]$ with $T = 2T_0$). In either situation we get a parameter set like the one needed to establish consistency.

If we do not have any way of knowing if there are "extra zeros" we should proceed as in Section 4.1.2 by plotting T against T minus the smallest zero in the criterion function.

These are of course only heuristic indications on how to proceed in the identification of a *good* parameter set and also how to prove consistency but the author is fairly confident that the conditions that are presented here, once established should be enough for consistency to hold.

4.5 On Establishing the Asymptotic Normality of the Estimator

If we where in a typical z-estimation setting, then Theorem 3.12 would give us the conditions under which we would have an asymptotically normal distributed estimator: a Lipschitz condition on the score function, the existence and finiteness of the second moment of each random variable we are working with when the core function is applied to them using the correct value for the parameter, the differenciability of the same quantity with respect to the parameter θ at θ_0 and that $\mathbb{P}_n\psi_{\hat{\theta}_0} = o_p(1/\sqrt{n})$. Our estimator is however not a z-estimator. While this would require us to derive a result specifically for our type of estimator, we can still, however, look at what the asymptotic variance of the estimator would be in case this result would be applicable. The asymptotic variance would be in that case $\frac{\int_0^{\theta_0}\lambda(t)dt+\int_{T-\theta_0}^{T}\lambda(t)dt}{(\lambda(\theta_0)-\lambda(T-\theta_0))^2}$ or

$$\frac{\int_0^{T-\tau}\lambda(t)dt + \int_\tau^T \lambda(t)dt}{(\lambda(T-\tau) - \lambda(\tau))^2} \tag{4.20}$$

This expression would only be valid if the derivative of the expectation of the score function was not null at θ_0. This will happen at least when T is a multiple of τ. So we would not expect our estimator to perform appropriately when a value for T is used that is a multiple of the period τ. Looking at this tentative variance of the estimator it also becomes clear that if we were to scale the intensity function by a factor of κ that the variance would be reduced by the same factor. So, even thought the scale of the intensity function would seem to play no role in terms of consistency or the asymptotic normality of the estimator it does influence the estimator's asymptotic variance.

If we look however at Figure 6.10 for a graph of the variance for a specific choice for the intensity function λ we see that the expression does seem to provide a good approximation for the actual variance of the estimator. This would suggest that even though we are in a different setting, it is still possible to translate these conditions into our situation or even that the same conditions are still enough for the result to hold.

5. DEFINITION OF THE ESTIMATOR FOR INDEPENDENT SAMPLE

In this chapter we will present the definition of our estimator for the period of a Non Homogeneous Poisson Process assuming we have a sample of independent processes. We will refer to this estimator as the "second version" of the estimator.

The first version of the estimator has a disadvantage: even though as we will see in Chapter 6 gives good results, it was obtained assuming we have a set of trajectories coming from a family of NHPP all sharing the same intensity function. In practise this means that we either have the same phenomenon replicating itself independently or we are somehow capable of identifying moments in time when the these phenomena reset themselves (in the stochastic sense). While it is possible for such situations to arise – say, data being generated independently by some common pseudo-random mechanism for example – we are however more interested in defining a more general estimator based solely on a trajectory from a single process.

The obvious difference in this case is exactly the fact that if we take a trajectory from a single process and partition it into sub-trajectories of length T and then apply the previous estimator, the criterion function does not converge. This follows from the fact that while using IID data we could make sure that for each value of the parameter, the quantities being averaged were also IID; now, the integrals we are comparing correspond to intervals that little have to do with some point where the process resets itself; not having thus, for a fixed value of the parameter, the same distribution, a convergence result will most likely require some constraint on the variance of the process or – and this is what is done here – the criterion function is modified so as to reduce its variance: this is achieved by working with the absolute value of the difference between the integrals considered in the previous chapter. We will see that this will reduce, for each value of θ, the variance enough so that we have convergence of the criterion function. As a consequence we will no longer be working with zeros of a criterion function (the criterion function will now be strictly non-negative) and will be looking instead for a value in the parameter set that minimises the criterion function. We shift now to an estimator that is quite similar to an M-estimator.

In this chapter we will start by defining another estimator for τ and looking at some of its properties.

5.1 Definition of the Estimator

Let us assume we have a NHPP $N(t), t \in [0, nT]$ with n going to infinity and that we will split the process in intervals of length T as:

$$N(t) = N_1(t)1_{\{t\in[0,T]\}} + N_2(t)1_{\{t\in[0,2T]\}} + \cdots + N_n(t)1_{\{t\in[0,nT]\}}$$

For $N_i(t)$ NHPP with intensity function $\lambda_i(t) = \lambda(t + (i-1)T)$ with $t \in [0, T]$.

Note that the collection $\{N_1, ..., N_n\}$ is not identically distributed but is mutually independent due to the processes' independent increments; each N_i is however a NHPP with an intensity function $\lambda_i(t)$ with period τ on the interval $[0, T]$. We will still define $D_i(t) := \overrightarrow{N}_i(t) - \overleftarrow{N}_i(t)$ as in Section 4.1 but we will also consider $S_i(t) := \overrightarrow{N}_i(t) + \overleftarrow{N}_i(t)$.

Considering the same criterion function as in Chapter 4 will not produce adequate results. To assure asymptotic convergence when working with a sample that is not identically distributed (for each t, the sample $\{D_i(t)\}$) one needs a condition on the variance of the elements of the sample

(Theorem 3.5). Estimating the criterion function (4.16) using a non identically distributed sample we can see that the function does not converge in general (Check Figure 5.1 for a comparison of estimates for (4.16) using an IID sample and a sample where we have only independence using the intensity function (5.1)).

$$\lambda(t) = 5 \sin\left(\frac{2\pi}{50} t\right) \tag{5.1}$$

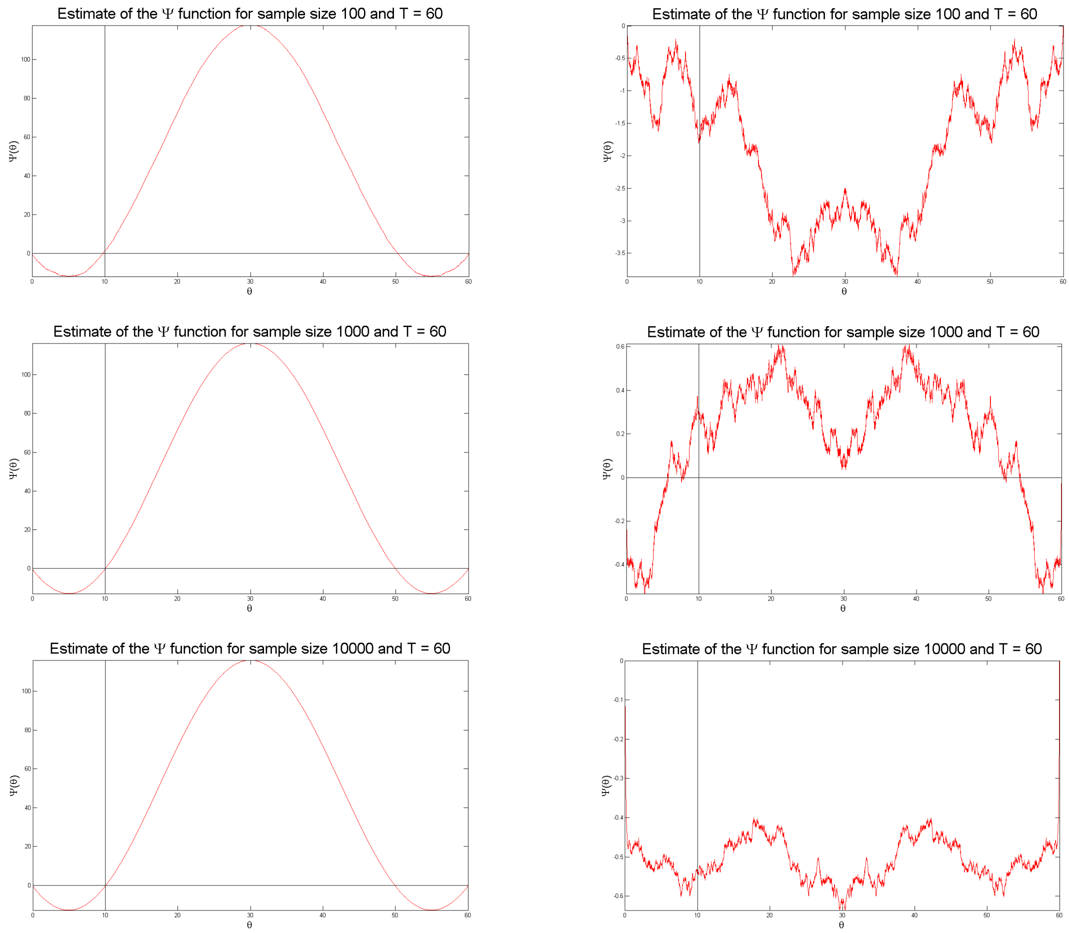


Fig. 5.1: Comparison of estimates for (4.16) using increasingly larger samples (rows) for IID sample and independent sample (columns).

By working for each t with the absolute value of the differences $D_i(t)$ the variance of the sample is reduced. By applying Jensen's Inequality (Theorem 3.14) we get that for any random variable X, $var(|X|) \leq var(X)$ as long as the necessary moments are defined. As such, working with a criterion function

$$\Psi_n(\theta) = \mathbb{P}_n \psi_\theta = \frac{1}{n} \sum_{i=1}^{n} |D_i(\theta)| \tag{5.2}$$

seems to be sufficient to assure convergence. This criterion function will then be strictly nonnegative and we therefore hope it retains enough similarities with an M-estimator so that we can estimate the period by looking at its local minima. A further modification was made to (5.2). Experimental results show that if we scale down the intensity function then we will be, at least to some extent, scaling down scaling down the criterion function. This entices problems on two levels: criterion functions for NHPP's whose intensity remains at low levels (comparatively) will result in less prominent local minima which are more difficult to identify in practise; also at least in M- and Z-estimation the speed at which the criterion function varies in a neighbourhood of the correct value for the parameter is related (inversely proportional even) to the asymptotic variance of the estimator. It is therefore desirable, if possible, to make sure that the impact of the scale of the intensity function affects the criterion function as little as possible since we have no control over the intensity function. It so turns out to be that by working with relative differences we can minimise the effects of the scale of the intensity function.

We will therefore be working with

$$\Psi_n(\theta) := \frac{\sqrt{\eta}}{\sqrt{n}} \frac{\sum_{i=1}^{n} |D_i(\theta)|}{\sqrt{\sum_{i=1}^{n} S_i(\theta)}} \tag{5.3}$$

with $\eta = \frac{1}{1-\frac{2}{\pi}}$. The sum in the denominator should be of order $o(n)$ and so the introduction of square root of this quantity (and the compensation using the factor $\sqrt{n}$) was not mandatory. This particular form, however, does allow us to write the criterion function as a ratio between means (or between a mean and the square root of a mean) and seems like a sensible choice. There is however another reason for this that will be show below; but before this we note that the constant η is, of course, innocuous, considering the use we will be giving this criterion function since it will not change the position of any local minima that the function might have. The quantity was however included to allow us to verify something related with the variance of the numerator of the criterion function.

We have mentioned that the motivation for working with the absolute value of the intensity function was to reduce the variance of the differences D_i and compensate for the fact they now do not share a common distribution. It is then interesting to look at what the variance of the absolute value of the difference between two Poisson distributed random variables is exactly. Jensen's inequality (Theorem 3.14) would give us only that $var(|D_i|) \leq var(D_i)$. It is interesting to see however that in our case the two mentioned variances are directly proportional.

Our first step was to simulate this variance. Figure 5.2 contains the simulation results for $var|N(\theta_N) - M(\theta_M)|$, i.e. the variance of the absolute value of the difference between a Poisson random variable with mean θ_N and one with mean θ_M.

The planes that we see in the graphs correspond to two cuts of the surface where θ_N and θ_M are proportional to each other. The graphic then suggests that as long as $\theta_N = \alpha\theta_M$ for some constant α, that we have $var|N(\theta_N) - M(\theta_M)|$ proportional to $\theta_N + \theta_M$. The most interesting situation should however be when $\theta_N = \theta_M$ since we know that this is the situation when θ takes the value θ_0.

The simulation results only suggest what was just mentioned the result would still have to be proved; we would have to establish that for N, M be two random variables Poisson distributed with mean t. Then

$$var|N - M| = 2\left(1 - \frac{2}{\pi}\right)t$$

We would then have $\eta := \frac{1}{1-\frac{2}{\pi}} > 1$.

We can of course look at both N and M like homogeneous Poisson processes with intensity 1 and so $N(t) \sim P(t)$ and $M(t) \sim P(t)$. The proof would have two steps: we would first prove that the variance $var|N(t) - M(t)|$ is of the form αt for some constant α and we would in a second stage determine the proportionality constant α.

Under the hypothesis that the process $|N(t) - M(t)|$ has independent increments, the result would be established as follows:

We need to show that for all $\beta \geq 0$ it holds that $var|N(\beta t) - M(\beta t)| = \beta var|N(t) - M(t)|$.

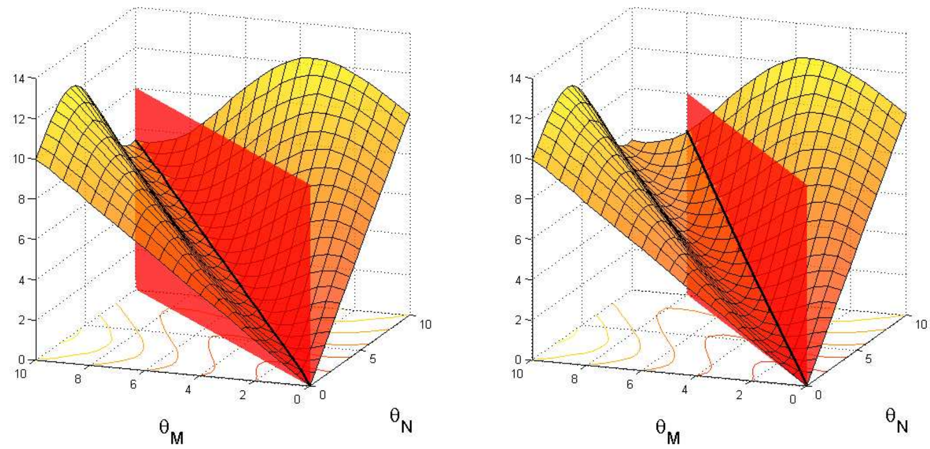


Fig. 5.2: Graph of simulated results for $var|N(\theta_N)-M(\theta_M)|$, the planes $\theta_N = \theta_M$ (left) and $\theta_N = 1.25\theta_M$ (right) and respective intersections.

Let us first prove this for $\beta \in \mathbb{N}$; we can decompose the process $|N(\beta t) - M(\beta t)|$ in increments of length t: $|N(\beta t) - M(\beta t)|$ is distributed like $|N(\beta t) - M(\beta t) - (N((\beta-1)t) - M((\beta-1)t))| + |N((\beta-1)t) - M((\beta-1)t) - (N((\beta-2)t) - M((\beta-2)t))| + \ldots + |N(t) - M(t) - (N(0) - M(0))|$;these increments are all distributed like $|N(t) - M(t)|$ and independent and so $var|N(\beta t) - M(\beta t)|$ is the sum of the variances of the increments, i.e. $\beta var|N(t) - M(t)|$.

Let us now prove the result for $\beta \in \mathbb{Q}^+$, i.e. take $\beta = \frac{a}{b}$ for $a, b \in \mathbb{N}$. We have $var|N(\beta t) - M(\beta t)| = a\ var|N(\frac{1}{b}t) - M(\frac{1}{b}t)|$ but now lets take t' such that $bt' = t$ then $var|N(t) - M(t)| = var|N(bt') - M(bt')| = b\ var|N(t') - M(t')|$ or $\frac{1}{b}var|N(t) - M(t)| = var|N(t') - M(t')| = var|N(\frac{1}{b}t) - M(\frac{1}{b}t)|$ so finally $var|N(\beta t) - M(\beta t)| = \beta var|N(t) - M(t)|$.

From a computational point of view this would be sufficient since we can only represent rational numbers when working with a computer; it is however quite simple in this case to make the step from rational coefficients to real ones. We already have that $var|N(t) - M(t)| = tvar|N(1) - M(1)| = \alpha t$, for $\alpha = var|N(1) - M(1)|/2$ and α rational; so let us take $\beta \in \mathbb{R}$ and consider any sequence $t_n \downarrow \beta$. We have $var|N(t_n) - M(t_n)| \downarrow var|N(\beta) - M(\beta)|$ since Poisson processes are right continuous and also $\alpha t_n \downarrow \alpha\beta$ and so, by the unicity of the limit of sequences of real numbers, $var|N(\beta) - M(\beta)| = \beta\alpha$ for any real β; we now proceed to determine the value of α.

If we consider the sequence $N_i \sim P(1)$ and $M_i \sim P(1)$ for $i \in \mathbb{N}$; we have that $N_i - M_i$ has mean zero and variance 2. Also, using the CLT

$$\sqrt{n}\sum_{i=1}^{n}(N_i - M_i) \xrightarrow{P} N(0,2)$$

and by the Continuous Mapping Theorem (Theorem 3.1)

$$|\frac{\sum_{i=1}^{n} N_i}{\sqrt{n}} - \frac{\sum_{i=1}^{n} M_i}{\sqrt{n}}| \xrightarrow{P} |N(0,2)|$$

And so, if the sequence $|\frac{\sum_{i=1}^{n} N_i}{\sqrt{n}} - \frac{\sum_{i=1}^{n} M_i}{\sqrt{n}}|^2$ is tight we should have by Theorem 3.13 that

$$\frac{1}{n} var|N(n) - M(n)| \to var|N(0,2)|$$

for $N(n) \sim P(n)$ and $M(\gamma n) \sim P(\gamma n)$. So using the established linearity,

$$var|N(1) - M(1)| = var|N(0,2)|$$

The random variable $|N(0,2)|$ is a half normal random variable and has variance

$$(2 - \frac{4}{\pi}) \tag{5.4}$$

So, we know that $var|N(n)-M(n)|$ is of the form αn and that $\frac{1}{n}var|N(n)-M(\gamma n)|$ converges to $(2-\frac{4}{\pi})$ then α is necessarily $2(1-\frac{2}{\pi})$.

We note that the result would hold regardless of the value of t; it would be inconvenient if this was not the case since in our case t will asymptotically take the value of the integral of the intensity function in the interval $[0, T-\tau]$: this would introduce another dependency on the value of T in the criterion function that we would probably not be able to take advantage of. There is however an advantage in relating the variances of $|D_i|$ and D_i since the variance of the latter is easy to compute since each D_i is simply the difference of two Poisson processes. We have thus found an estimator for the variance of $|D_i|$. We then see that for $\theta = \theta_0$ the denominator in our criterion function simply corresponds to an estimate of the square root of the variance of the estimator. As for when the estimates for the areas being compared don't match (which should happen if $\theta \neq \theta_0$), then the quantity in the denominator should be smaller than the square root of the variance of the numerator. In Figure 5.3 we show the comparison.

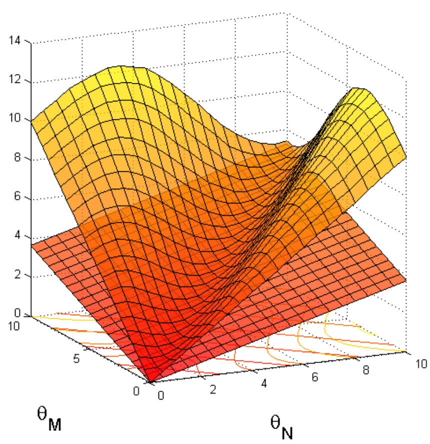


Fig. 5.3: Comparison of the simulated variance of the absolute value of the difference of two Poisson random variables as a function of their parameters (surface) and a plane containing the variance over $\theta_M = \theta_M$ and perpendicular to this same plane.

We can see that indeed the expression that was derived coincides with the simulated variance of the absolute value of the difference on the plane $\theta_N = \theta_M$ and that this expression is always below the simulated variance outside the same plane.

Note that Ψ cannot be defined as in (4.16), meaning as an expectation, since for fixed θ, the $D_i(\theta)$ random variables do not share a common distribution. We will simply be defining $\Psi(\theta)$ as, for each $\theta[0, T]$, the limit of $\Psi_n(\theta)$. More needs to be said about this, namely if $\Psi(\theta)$ is indeed defined, but this will be done in the following sections.

Our estimator will be obtained from a sequence of estimators $\hat{\theta}_n$ that (nearly) minimises the equations (5.3) an is thus an similar to an M-estimator. And is defined as

$$\hat{\theta}_n := \theta' \in \Theta : \Psi_n(\theta') = \inf_{\theta \in \Theta} \Psi_n(\theta) \tag{5.5}$$

and finally

$$\hat{\tau} = T - \hat{\theta}_n \tag{5.6}$$

We further note that since we take $\hat{\theta}_n$ to be an infimum it will always be defined if we take Θ as a closed interval, for example.

Some properties need to be derived for the criterion function.

5.2 Properties of the Estimator

In this section we will be presenting some properties of the estimator and the criterion function. We aim at obtaining our estimate based on the identification of a local minima in estimates of some function Ψ that we know has a local minima at the correct value of the parameter $\theta_0 = T - \tau$. That this is a sensible idea and that it can actually be done still needs, however, to be justified.

A first step would be to establish that for every $\theta \in \Theta \subset [0, T]$ the sequence of criterion functions $\Psi_n(\theta)$ converges (to some function $\Psi(\theta)$).

Proposition 5.1. There exists a function $\Psi(\theta)$ such that for every $\theta \in \Theta \subset [0, T]$, for Θ compact, the sequence of random functions $\Psi_n(\theta)$ converges a.s. to $\Psi(\theta)$.

Proof. What needs to be proved here is not so much the existence of Ψ but that for all θ the sequence $\Psi_n(\theta)$ indeed converges since then we can just take $\Psi(\theta)$ as the sequence's almost sure limit. Let us then take $\theta \in \Theta$;we can rewrite $\Psi_n(\theta)$ as

$$\Psi_n(\theta) = \sqrt{\eta}\frac{\frac{1}{n}\sum_{i=1}^n |D_i(\theta)|}{\sqrt{\frac{1}{n}\sum_{i=1}^n S_i(\theta)}}$$

The limit we are interested can therefore be obtained as the ratio of the limits of the numerator and denominator of this expression. Since both the numerator and (the square of) the denominator are means we will use Theorem 3.5, the SLLN for independent samples, to prove that these means converge – for this, it suffices to prove that the variance of the random variables being summed is no more than $o_p(n^2)$.

Let us take $k \in \mathbb{N}$ such that $T \leq k\tau$ (say $k = \lceil \frac{T}{\tau} \rceil$); we have that for all i, $|D_i(\theta)| \leq S_i(\theta) \leq 2N(k\tau)$ with N our initial, unpartitioned, NHPP which has finite expectation. We thus have

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n var|D_i(\theta)|}{n^2} \leq \lim_{n\to\infty} \frac{\sum_{i=1}^n var[S_i(\theta)]}{n^2} \leq \lim_{n\to\infty} \frac{\sum_{i=1}^n var[2N(k\tau)]}{n^2} = \lim_{n\to\infty} 4\frac{var[N(k\tau)]}{n} = 0 < \infty$$

which hold a.s..

We thus have $\frac{1}{n}\sum_{i=1}^n |D_i(\theta)| - \mathbb{E}|D_i(\theta)|$ and $\frac{1}{n}\sum_{i=1}^n S_i(\theta) - \mathbb{E}S_i(\theta)$ both converge, a.s. to zero. The issue is now whether $\frac{1}{n}\sum_{i=1}^n \mathbb{E}|D_i(\theta)|$ and $\frac{1}{n}\sum_{i=1}^n \mathbb{E}S_i(\theta)$ converge. We of course assume these quantities have different distributions, but under the assumption that both τ and T are rationals this is not true: there necessarily exists a constant, say m, corresponding to the number of different distributions that these random variables can take: it will be the least common multiple between τ and T; we therefore have that elements in the sample such that the difference between their indexes is a multiple of m have the same distribution. Then trivially $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|D_i(\theta)|$ converges to $\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}|D_i(\theta)|$ and $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}S_i(\theta)$ converges to $\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}S_i(\theta)$. □

We now know that Ψ is well defined and not random. We now see that our setting of different distributions is not was bad as it could be.

Another important property to establish for the estimator is continuity. This is important from both a theoretical point of view: if θ_0 is a discontinuity point we might have problems defining the estimator's variance since, as we have already seen in the previous chapter, it is related with the derivatives at this point; and a practical point of view: we will always be working with a discrete set of points when estimating the criterion function, and so we might be considering only points in a neighbourhood of θ_0 that due to discontinuities might have quite different values from the function immediately around θ_0. We will not be establishing this result but do conjecture however that for $\Theta = [0, T/2]$ we have that $\Psi(\theta)$ is a.s. continuous.

Proposition 5.2. Let Θ a compact subset of $[0, T/2]$ then we have $\Psi(\theta)$ a.s. bounded over Θ.

Proof. Considering that Ψ can be written as the ratio between $\sum_{i=1}^{\infty}\mathbb{E}|D_i(\theta)|$ and the square root of $\sum_{i=1}^{\infty}\mathbb{E}S_i(\theta)$ then this quantity is trivially bounded (over θ) by $\sqrt{2\eta\mathbb{E}N(\lceil\frac{T}{\tau}\rceil\tau)}$. □

The criterion function, even though defined based on differently distributed quantities, can still be proved to be a.s. bounded and to convergence to a finite limit; we further conjecture it to be continuous. It does then still contain analogous properties to the ones we saw for the first version of the estimator and certainly to those we would expect from an actual M-estimator. In the next sections we will shift our focus to the estimator itself.

5.3 On Establishing the Consistency of the Estimator

Since we will be working with a compact parameter set, the positivity and a.s. finiteness of Ψ_n and Ψ assure the existence of local minima. The issue then becomes if the local minima of Ψ_n indeed converge in probability to the local minimum of Ψ and if Ψ a.s. has a local minimum at $\theta_0 = T - \tau$.

We start by establishing the latter. Note that while it might seem intuitive that given two random variables, say X with null expectation and Y with positive expectation that $\mathbb{E}|X| \leq \mathbb{E}|Y|$ this is not true in general; however this will hold in our case.

Proposition 5.3. Let $D(t) = N(\overrightarrow{\lambda}(t)) - N(\overleftarrow{\lambda}(t))$ then $\mathbb{E}[|D(t)|]$ has a local minimum at a point t' (which we assume is unique) such that $\overrightarrow{\lambda}(t') = \overleftarrow{\lambda}(t')$ if we consider also[1] that both $\overrightarrow{\lambda}$ and $\overleftarrow{\lambda}$ are continuous and that $\overrightarrow{\lambda}(t) - \overleftarrow{\lambda}(t)$ changes sign at some point t' (unique).

As a consequence $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|D_i|]$ has a local minimum at a point such that $\overrightarrow{\lambda} = \overleftarrow{\lambda}$.

Proof. Note that to establish the second part of the result it is then only necessary to establish the result for an individual term; also, if we take some $t_0 > t'$ we can assume without loss of generality that $\overrightarrow{\lambda}(t_0) > \overleftarrow{\lambda}(t_0)$ and due to the presence absolute value it suffices to establish the result for this case.

We then have $\lambda_2 = \overleftarrow{\lambda}(t_0) \leq \overleftarrow{\lambda}(t') \leq \overrightarrow{\lambda}(t_0) = \lambda_1$. Using the merging and splitting properties of Poisson random variables we have[2]: $|N(\lambda_1) - N(\lambda_2)|$ is distributed like $|N(\lambda_1) + N(\lambda_1 - \lambda(t')) - N(\lambda_2) - N(\lambda(t') - \lambda_2) + N(\lambda(t') - \lambda_2)|$ which in turn is distributed like $|N(\lambda(t')) - N(\lambda(t')) + N(\lambda_1 - \lambda_2)|$ and so $\mathbb{E}|N(\lambda_1) - N(\lambda_2)| = \mathbb{E}|N(\lambda(t')) - N(\lambda(t')) + N(\lambda_1 - \lambda_2)|$; we only establish that $\mathbb{E}|N(\lambda(t')) - N(\lambda(t')) + N(\lambda_1 - \lambda_2)| \geq \mathbb{E}|N(\lambda(t')) - N(\lambda(t'))|$.

[1] Note that this is simply the setup for our situation.

[2] Whenever we use in the same expression $N(\dot{})$ twice we mean by this that each is a Poisson random variable independent from all other we are considering.

Let us abbreviate the notation and refer to $N(\lambda(t')) - N(\lambda(t'))$ as Y and to $N(\lambda_1 - \lambda_2)$ as X;we then hope to establish that $\mathbb{E}|X + Y| \geq \mathbb{E}|Y|$.

We have $\mathbb{E}|X+Y| = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\mathbb{P}(X = j)\mathbb{P}(|Y| = i-j)$ which can be rewritten as a conditional expectation $\mathbb{E}|X + Y| = \sum_{j=0}^{\infty}\mathbb{P}(X = j)\mathbb{E}[|X + Y|\,|X = j]$; so, if for every $j \in \mathbb{N}$ we have that $\mathbb{E}|j + Y| \geq \mathbb{E}|Y|$ the result follows. This will be a consequence of two facts: the random variable Y is such that $\mathbb{P}(Y = k) = \mathbb{P}(Y = -k)$ and $\mathbb{P}(Y = k)$ decreases with k. The first is an immediate consequence of Y being the difference of two independent random variables with the same distribution; the second is a consequence of Corollary 3.1 and the fact that the modified Bessel function $I_n(z)$ decreases, for fixed z with n.

Let us consider the sequence $a_n = \mathbb{P}(Y = n)$ and $b_0 = 0, b_{2n-1} = b_{2n} = n$. We then have $\mathbb{P}(|Y| = n) = 2a_n, n > 0$ and so from what was said before $a_{b_n}, n > 0$ is decreasing; b_n is on the other hand decreasing; we thus have that

$$\mathbb{E}|Y| = \sum_{i=1}^{\infty} b_i a_{b_i}$$

We also have that for any $j \in \mathbb{N}$, $\mathbb{P}(|Y + j| = n) = a_{n-j} + a_{-n-j}$ and so, if we write

$$\mathbb{E}|Y + j| = \sum_{i=1}^{\infty} b_i c_{b_i}$$

then c_n is a reordering of a_n; the result follows then from the rearranging inequality (Theorem 3.16). □

This only proves that the numerator of our criterion function is locally minimised at $T - \tau$ the denominator is however a function that is in mean an increasing function; it might so happen that the numerator might distort thus the local minimum into higher values for the parameter; this should in principle not happen since the normalisation factor that we use changes linearly for linear growth of the parameter. Another heuristic justification for this is that the numerator is $o(n)$ and the denominator of order $o(\sqrt{n})$ and so, the numerator should dominate in that the denominator should be roughly constant when compared to the numerator. we are however at fault here by not establishing the we indeed have a local minimum at $T - \tau$ for the full criterion function – we will however proceed with the assumption that this is indeed the local minimum for the criterion function.

Our next step would be to establish the consistency of an estimator based on this criterion function. This is surely not easier for the second version of the estimator than for the first version, the main difference being the presence of differently distributed random variables. The procedure should however be similar: if we are able to construct a parameter set that contains $T - \tau$ and no other points of the form $T - k\tau$ for $k \in \mathbb{N}$ and we do have an isolated local minimum $T - \tau$ then it should suffice to establish a uniform convergence (over θ) of the criterion functions Ψ_n to Ψ. Under these assumptions, the proof should go as in the IID case, as the difficulty of having differently distributed random variables should only translate into it being more difficult to establish the uniform convergence result.

We are not totally at a loss here, however. The fact that we will be partitioning a trajectory means that we are, in each element of the partitioned trajectory comparing (sub-)trajectories generated from different pairs of regions of the intensity function; this *necessarily* makes it more difficult for these event counts to coincide at unwanted points (meaning points other than $T - \tau$): the constraints on the shape of the intensity function that where derived for these unwanted points would now be several times folded; this suggests that identifyability should be less of a problem using this version of the estimator.

5.4 On Establishing the Asymptotic Normality of the Estimator

The most important step in establishing the asymptotic normality of Z- and M-estimators consists in proving that $\sqrt{n}\Psi_n(\theta_0) - \sqrt{n}\Psi(\theta_0)$ is asymptotically normal. The rest follows from developing $\Psi_n(\theta_n)$ as a Taylor series around θ_0 and proving that the term corresponding to the second derivative of Ψ_n is of a small order: we would in this way directly relate the difference between the sequence of criterion function and its limit and the difference between the estimates and the correct value for the parameter. This should then be an indication of a sensible path to follow to establish normality for this version of the estimator. That $\sqrt{n}\Psi_n(\theta_0) - \sqrt{n}\Psi(\theta_0)$ is asymptotically normal is not an unfounded idea: it should simplify to simply a centered and normalised sum of independent random variables. Lyapunov's CLT (Theorem 3.7) or Lindenberg's CLT (Theorem 3.8) then give us conditions under which the limit of these quantities is normaly distributed.

6. IMPLEMENTATION AND RESULTS

In this chapter we will be presenting some experimental results obtained using our estimators. We will begin by talking a bit about the algorithmic implementation of the estimators and how the data we worked with was obtained and show the intensity function that we will be working with throughout most of this chapter; we will then proceed to using the estimators in the fashion one would when given Poisson data and asked to estimated the period in the data; after this we proceed to repeatedly estimate the quantities we work with to see how they behave in average and thus obtain more stable results; after this we experiment with changing the average number of events per cycle to further test the estimators – these tests will focus on estimation under a low number of events per cycle and short trajectories; we also compare our estimators with the ones found in the literature; and finally take a quick look at results obtained for other intensity functions.

6.1 Implementation Issues

In this section we will present aspects relative to the implementation of the algorithm.

It would be difficult to find, in practise, phenomena where the real value of the period is know; even if one such phenomenon was know there would certainly be limitations on the amount of data available. The experimentations presented here were then done using generated data. The Poisson data was simulated based on a chosen intensity function and then used to estimate the period. This way it would be possible to evaluate the actual quality of the estimate. We present now the algorithm (Algorithm 1) used to simulate the NHPP data.

Algorithm 1 Generation of a trajectory of a NHPP[6, p.84]

Require: Length of trajectory, T
Require: Intensity function $\lambda(t)$
Require: Upper bound λ for $\lambda(t)$
Ensure: Vector of event times S

```
t = 0;
i = 0;
while t ≤ T do
   Generate random number U;
   t = t − (1/λ) log(U);
   if U ≤ λ(t)/λ then
      i=i+1;
      S(i) = t;
   end if
end while
```

As for the algorithm used to perform the estimation itself, it works by partitioning the interval $[0, T/2]$ in which we are working into a grid[1] $t_0 = 0, t_1, ..., t_r = T/2$ and keeping track of the

[1] In our case we worked with a grid spacing of 0.01 for the considered interval.

cumulative difference – in the case of the first version of the estimator – and the cumulative absolute difference and cumulative sum– for the second version of the estimator – of the number of events in $[0, t_i]$ and $[T - t_i, T]$; these quantities, and the total number of trajectories of length T that we have are sufficient to obtain an estimate for Ψ_n; we then have one function to search for a zero and one to search for a local minimum that we apply to Ψ_n to obtain an estimate for $T - \tau$ which is then deducted from the value of T to produce our estimate for the period. We reproduce in Algorithm 2 the structure of the algorithm.

Algorithm 2 Estimation of the period

Require: Length of trajectory, T
Require: Number of subintervals of length T, n
Require: Version of the estimator, "version", taking the values 1 or 2
"times" $= \{0, 0.01, 0.02, \ldots, T/2 - 0.01, T/2\}$
Define "differences" as a null vector with the same number of elements as "times"
Define "cumulatives" as a null vector with the same number of elements as "times"
for i is 1 to n **do**
  if version = 1 **then**
    data = $NHPP(T, 0)$
    differences = differences + diffs(data, times, T)
  end if
  if version = 2 **then**
    data = NHPP(T, T(i-1))
    differences = differences + diffs(data, times, T)
    cumulatimes = cumulatives + cumuls(data, times, T)
  end if
end for
if version = 1 **then**
  estimate = T - get_zero(differences/n)
end if
if version = 2 **then**
  estimate = T - get_min(differences/($\sqrt{n \times \text{cumulatives} \times (1 - 2/\pi)}$))
end if

Some notes about the algorithm; NHPP is the function seen in Algorithm 1 with the only difference that we have a second argument corresponding to the starting point for the intensity function (for when we are partitioning the trajectory of the process); the functions "diffs" and "cumuls" return for fixed T and given a trajectory of that length ("data") the difference and sum (respectively) between the number of events in $[0, t_i]$ and $[T - t_i, T]$ for each event time t_i in "times"; "get_zero" and "get_min" return the zero and minimum (respectively) for the quantities passed as argument and return one such point only if one is to be found.

As for the intensity function that was used itself, due to the difficulty (or even impossibility) of choosing a set of families of intensity functions that would be in any way representative of all periodical functions we might want to work with, we present here results for the main tests that we will be performing for only the following family of intensity functions:

$$\lambda(t) = \frac{\rho}{5}\left(\cos\left(\frac{2\pi}{\tau}t\right) + \cos\left(2\frac{2\pi}{\tau}t\right) + \cos\left(3\frac{2\pi}{\tau}t\right) + 2\right) \tag{6.1}$$

Note that (6.1) has period τ and maximum ρ. In [15] this intensity function proved to be the most challenging one for both the periodogram and the several versions of the estimator presented by the authors even when considering a quite ample range for an estimate to be considered "successful" – deviations smaller than 10% with respect to the true value of the period were considered successful estimations. Unless otherwise mentioned we will be working with $\tau = 50$ and $\rho = 10$. Note also that the expected number of events per period is given by

$$\frac{2\rho}{5}\tau \tag{6.2}$$

which in our case corresponds to 200 events being expected per our period of 50 time units.

6.2 Estimation Approach

In this section we will be presenting the estimation procedure as one would expect it to be used when confronted with real data. We begin with the first version of the estimator. It is important to do this since for a successful utilisation of the estimator, a good choice for T is crucial; so our main objective for this section is to identify one such "good" value.

6.2.1 First version of the estimator

The first step in the analysis would be to apply our estimator to the data for different values of T and examine the plot of the estimates versus the respective values of T. We considered a sample of size 1000 for each value of $T = 1, 2, ..., 110$. We tried to use throughout this chapter the same sample size for all of the tests so as to better allow comparisons; in practise, however, this graph can be constructed using relatively small samples since it supposed to be used solely as a graphical aid: while using larger samples will obviously give us more precision, it should not severely affect the shape of the graph since the variability that the estimates might present should always be of a small order when compared to the jumps between steps. We also mention that the use of large sample is meant is used to give an idea of the estimators asymptotic behaviour.

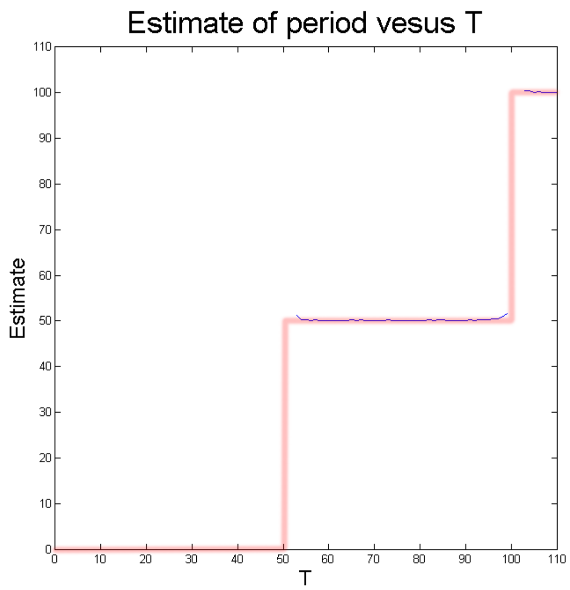


Fig. 6.1: Estimates for τ as a function of T.

We can clearly see the staircase function shape in Figure 6.1. We would stop trying to increase the value of T when a second height for the plateaux is reached and we have steps of the same length and height. The implementation of the estimator, as mentioned before, does not return a value when the resulting criterion function does not have a zero – the values of T corresponding to these situation correspond to the interruptions in the graph. This can happen in two situations: when T is less than the period and eventually when T is a value close to the period or its multiples. Other than this the functions is exactly what would be expected: a staircase functions with steps with length the same length and height (which should correspond to the value of the period).

This function should therefore allow us to identify "safe" choices for T: these should belong to the interval where the function first becomes constant at a value other than 0 – roughly 50 to 100 in this case. We show in Figure 6.2 the criterion function obtained for different choices of T

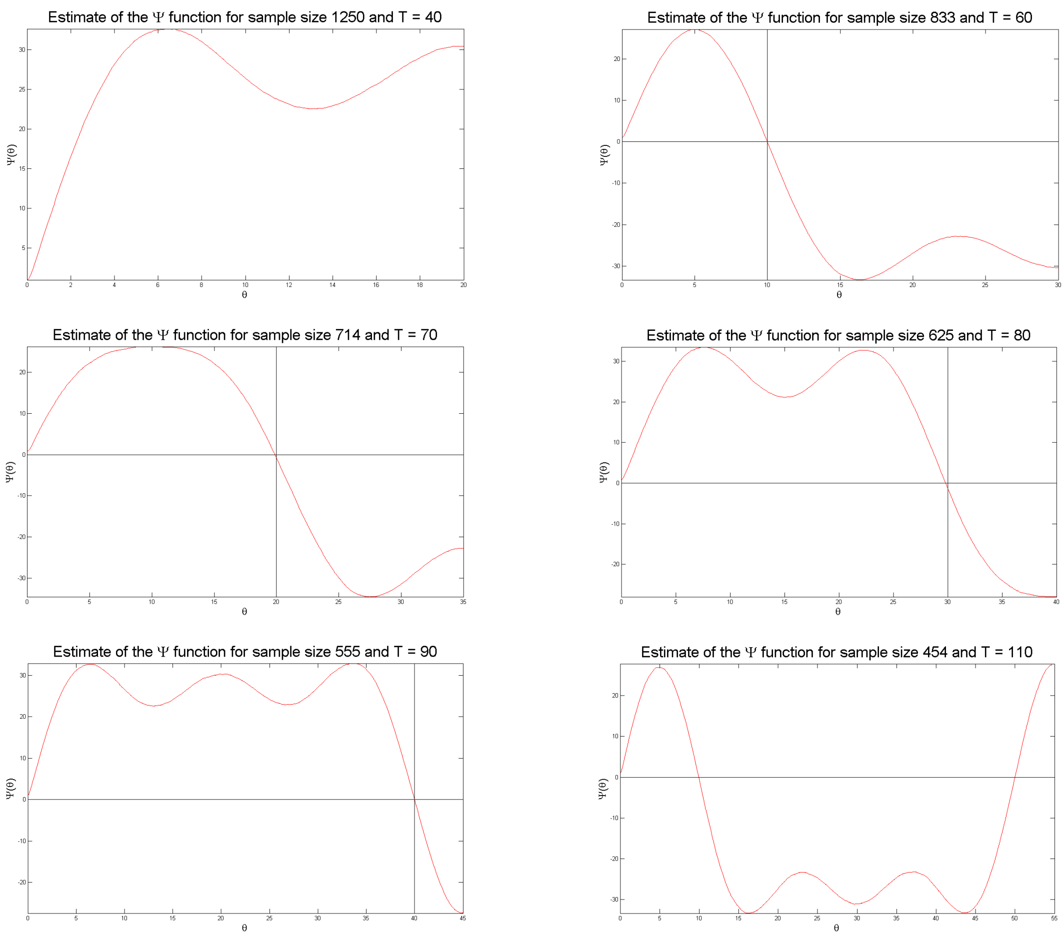


Fig. 6.2: Estimates of the Ψ function for T is 40, 60, 70, 80, 90, 110 (left to right and top to bottom).

As expected for $T = 40$, a value bellow the period, the criterion function has no zeros (except for $\theta = 0$ of course); for $T = 110$ we get two zeros – each one corresponding to the difference to one of the multiples of the period; For $T = 60, 70, 80, 90$ we get a unique zero at close to the correct position. Table 6.1 summarises the obtained estimates and respective relative deviations.

T	40	60	70	80	90	110
Estimate	–	50.005	50.135	50.185	49.995	100.055
Deviation	–	0.01%	0.27%	0.37%	-0.01%	100.11%(0.055%^{2})

Tab. 6.1: Estimates for the period and respective relative deviations for different values of T.

The next step in the analysis would be to actually choose a value of T to use; in principle, T should be chosen small since this will allow obtaining a larger sample to work with. Looking at the expressions derived for the asymptotic variance of the estimator (4.20), while choosing small T immediately reduces the numerator, unless we have sufficient knowledge about λ to identify T such that the denominator is as large was can be we cannot find an optimal choice for T.

Before we do focus on one value of T thought we will show in the next section similar results to those presented in this section but now for the second version of the estimator; further analysis of the first version of the estimator continues however in Section 6.3.

6.2.2 Second version of the estimator

We will repeat in the current section the tests made in Section 6.2.1 now for the second version of the estimator. We begin by presenting, for a sample size of 1000 and for $T = 1, 2, ..., 110$ the values returned by this version of the estimator.

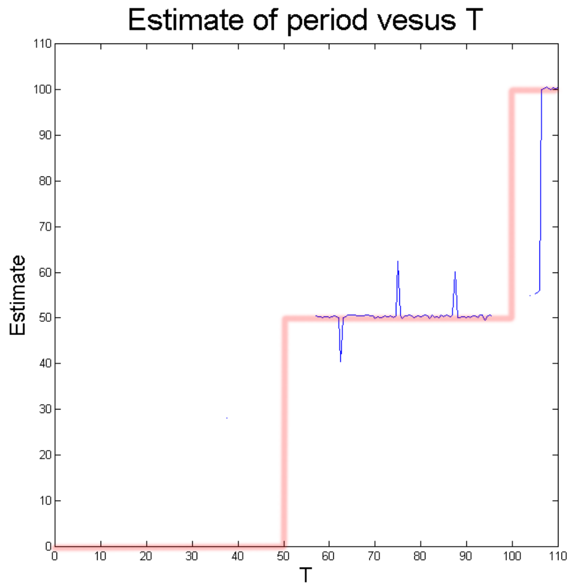


Fig. 6.3: Estimates for τ as a function of T.

We can observe no major difference between this graph and the one presented in Figure 6.1 except for some deviation close to multiples of the period and three values for T around which there seems to be some instability – this is however clearly visible. We now present (Figure 6.4) for $T = 40, 60, 70, 80, 90, 110$ our estimates for the Ψ function.

We are now identifying the estimate as a point corresponding to a local minimum in the criterion function. As we can see, there are, in this case, several local minima, for the values of T for which we have the peaks; this strengthens the notion that a sensible choice for T must be preceded by looking at a graph like Figure 6.3. We do see however that the difference to the actual multiple is the value of the parameter associated with the local minimum with the lowest value of Ψ (Check the estimate of Ψ for $T = 90$).

T	40	60	70	80	90	110
Estimate	–	49.90	50.33	49.95	49.76	100.04
Deviation	–	0.30%	0.62%	0.20%	0.04%	100.08%(0.04%[3])

Tab. 6.2: Estimates for the period and respective relative deviations for different values of T.

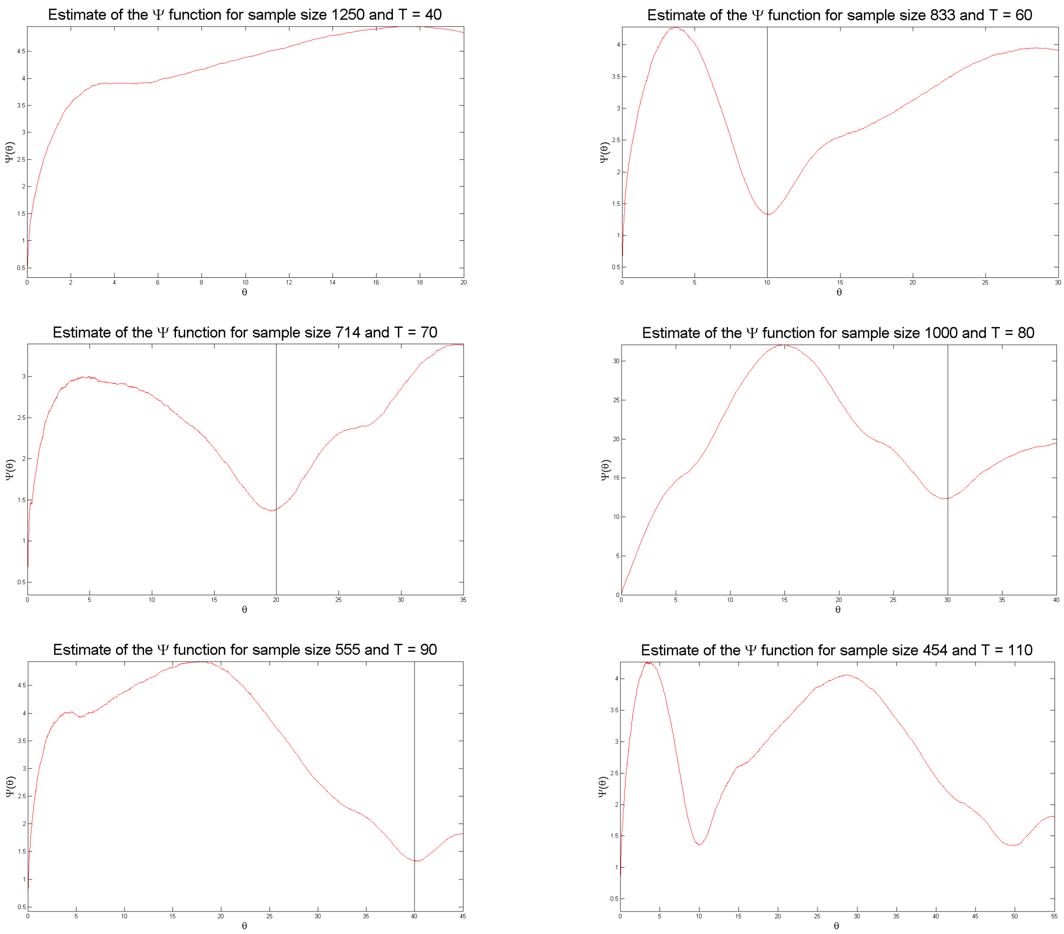


Fig. 6.4: Estimates of the Ψ function for T is 40, 60, 70, 80, 90 and 110 (left to right and top to bottom).

6.3 Repeated Estimation

We will now for both estimators do a repeated estimation of period to have more stable results and thus an idea of how the estimator behaves in general. It will also give us an opportunity of comparing the obtained variance for the estimates and the expression for the variance obtained in (4.20) and test the normality of the obtained estimates.

6.3.1 First version of the estimator

We start by repeating the estimation of the period, with different data, 250 times and average the results, always considering a sample size of 1000 for each estimation. We will be focussing on $T = 55, 60, 65, 70, 75, 80, 85, 90, 95$. We start by presenting pointwise averages for the several criterion function that we obtain to have an idea of its asymptotic shape.

As we can see (Figure 6.5) the shape of the average of the criterion function is quite similar to the shape of the individual estimates for the same quantity (check Figure 6.2). In all six cases we have a zero at $\theta = T - \tau$ besides the expected zero at $\theta = 0$. We also note that we are only representing the criterion function on the respective $[0, T/2]$ intervals since over $[0, T]$ we have symmetry.

Another interesting quantity to look at is the pointwise variance of Ψ_{1000} (in our case) – this is reproduced in Figure 6.6. The graphs in Figure 6.6 show the pointwise variance of our estimates of Ψ for the indicated values of T as a function of θ. The behaviour that these graphs present – with the variance showing a increasing trend with θ that seems, in this case, to be linear – was expected: we are representing quantities which are an estimate of the difference between Poisson random variables whose parameters increase with θ, immediately implying that the variance of these quantities increases with said θ. The type of growth that we see in general should of course depend on the obvious factors: the function λ itself and the value for T.

Next, we look at the evolution of the average of the estimates as a function of the sample size (Figure 6.7). In Figure 6.7 we see that for practically every choice of T the average result produced by the estimator seems to behave as expected converging to τ (50 in our case) – we look only at the estimates in the fixed range $[49.8, 50.2]$ since virtually all estimates fall within this range and also to allow for a more homogeneous view of the several graphs. We note that the actual range for the mean estimates was approximately $[49, 50.4]$: a maximal absolute relative difference of 2% across all sample sizes and all values of T.

We see that due to the proximity of the zero to the endpoint of the parameter set when $T = 95$ (and thus $\theta_0 = 5$) our algorithm is not always capable of identifying the zero of the criterion function when a low sample size is used; we note however that the absence of an estimate is only verified for this particular choice of T for $n < 75$ (approximately) and that it simply means that at least one of the estimates of the 250 used to compute the mean results was invalid. Other than this there does not seem to be much difference between the several graphs.

We also present the variance of the estimates as a function of the sample size (Figure 6.8). We can see in Figure 6.8 that all cases, as expected, the variance of the estimates produced by the estimator is reduced as the sample size is increased regardless of or choice for T. The logarithm of the variance seems also to be roughly proportional to $1/\sqrt{n}$; this behaviour seems to hold regardless of T and the difference seems to be essentially the range of values that the variance takes.

We further proceed by presenting for each value of T an histogram of the final estimates for the period (Figure 6.9).

All of the histograms seem to be centred at roughly 50 and the shape does not change much – the number of counts decreases as we step further away from the bin containing 50; the major difference between the graphs seems to be some asymmetry in some of them.

We performed Kolmogorov-Smirnov normality tests for these estimates. Table 6.3 contains the p-values for the tests as well as the obtained average estimates, variance, asymptotic variance (derived in Section 4.5) and kurtosis. We notice in Table 6.3 that in all cases we do not reject (at say 5%) the hypothesis that the estimates produces by a sample of size 1000 are normally

T	Average Estimate	Relative Deviation	Variance	Asymptotic Variance (4.20)	Kurtosis	p-Value
55	50.0005	0.0010%	0.0040	0.0044	3.1863	0.2559
60	49.9999	-0.0001%	0.0014	0.0015	2.7707	0.0745
65	50.0025	0.0051%	0.0030	0.0031	3.3818	0.6262
70	50.0039	0.0078%	0.0046	0.0043	2.6688	0.4381
75	50.0015	0.0031%	0.0036	0.0031	3.0262	0.2639
80	50.0117	0.0234%	0.0062	0.0056	2.9275	0.3806
85	50.0021	0.0042%	0.0074	0.0061	2.3917	0.1744
90	49.9994	-0.0012%	0.0037	0.0039	3.0219	0.2339
95	50.0074	0.0149%	0.0170	0.0165	2.5680	0.4499

Tab. 6.3: Summary of the obtained results for the first version of the estimator.

distributed. Also, in all cases our average estimate is very close the true value of the parameter. We can also see that the variances of our repeated estimates are quite close to those obtained in (4.20): in Figure 6.10 we plotted the logarithm of (4.20) against the logarithm of the obtained variances. The variance converges to infinity as T approaches a multiple of the period; for this reason we look only a smaller range for the (logarithm of the) variance.

For our choice of intensity function we see that the two quantities (meaning the proposed asymptotic variance and the estimates for the variance obtained by repeated estimation of the period) fit quite well. We indeed see how the variance of our estimator is quite high for choices of T close to multiples of the period; away from these points however the behaviour depends solely on the intensity function: we see and increasing trend on the variance as T increases but this trend is anything but monotone. We note that within the considered interval $T = 60$ seems to be the most appropriate choice when it comes to minimising the variance; this choice (being a low value for $T > \tau$) would also assure that we maximise the number of elements in our sample.

We thus see that not only does the estimator perform quite well but its results also fits nicely to what we expect from analysing the theoretical model.

When we further look into the results produced by the estimator we will therefore be considering $T = 60$; but, before we do so, we will first look in the following section into the behaviour of the second version of the estimator by repeating for it what was presented in the current section. Further analysis of the first version of the estimator continues in Section 6.4.1.

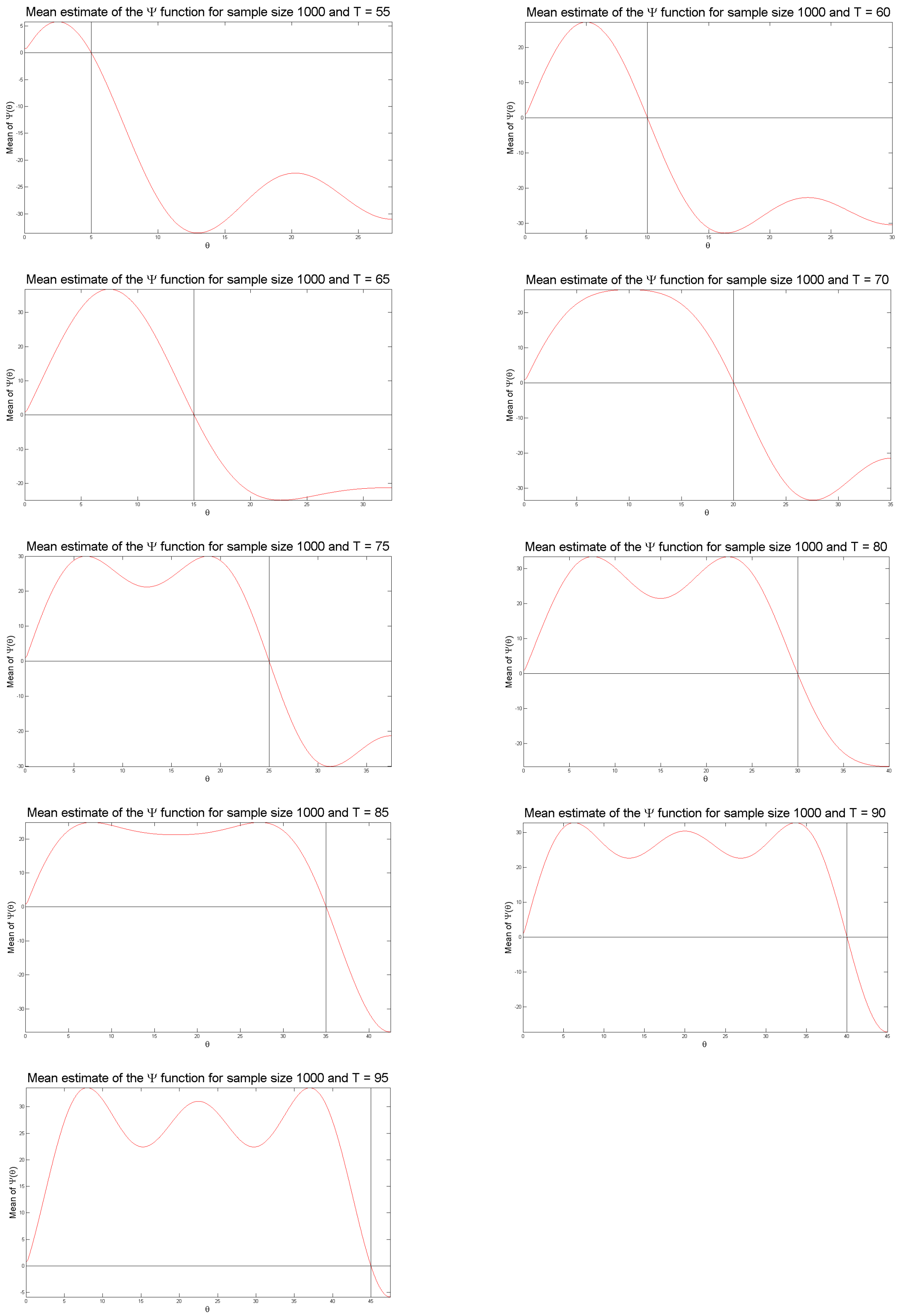


Fig. 6.5: Pointwise mean of the estimates of the Ψ function for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

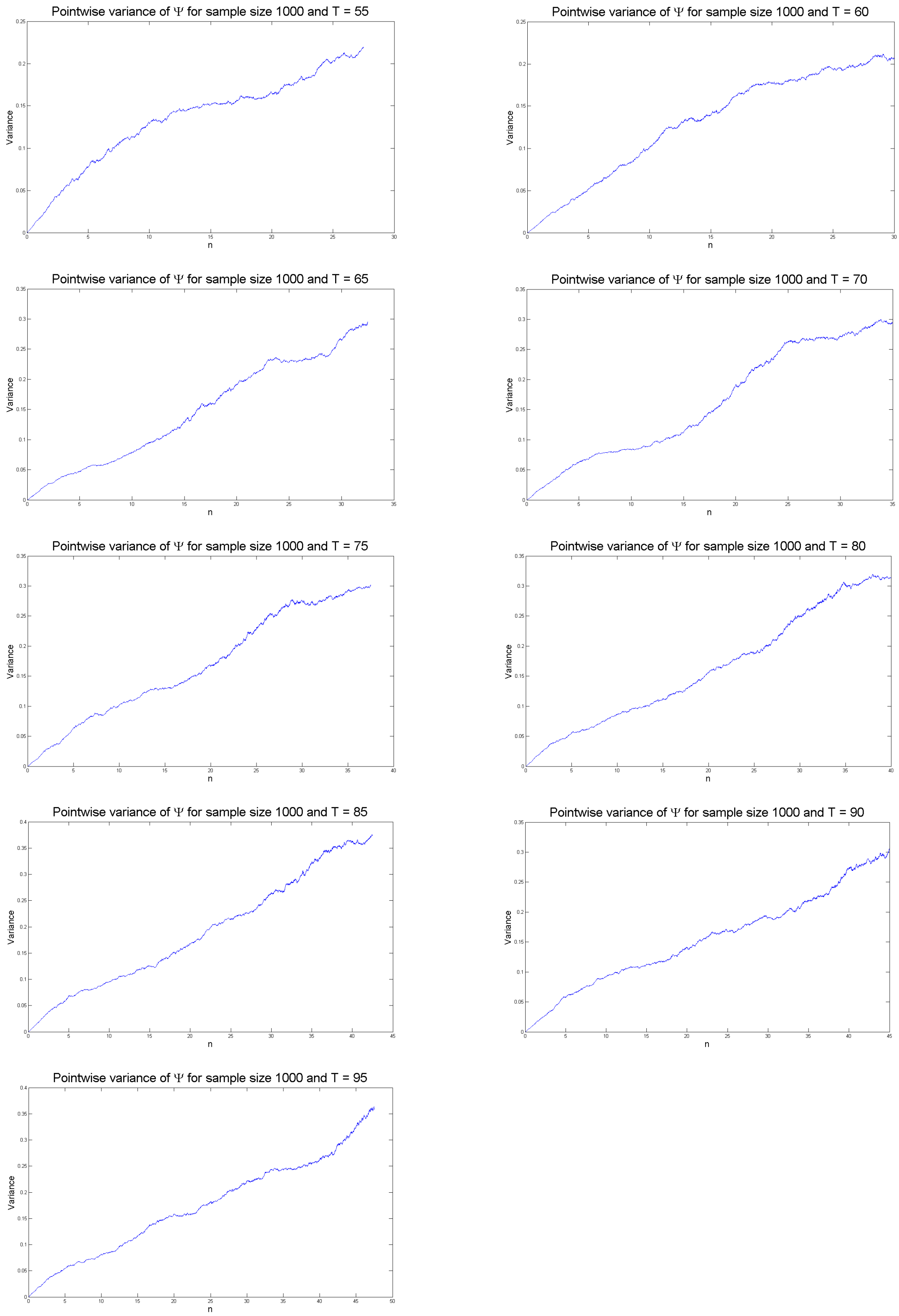


Fig. 6.6: The variance of the estimates of the Ψ function for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

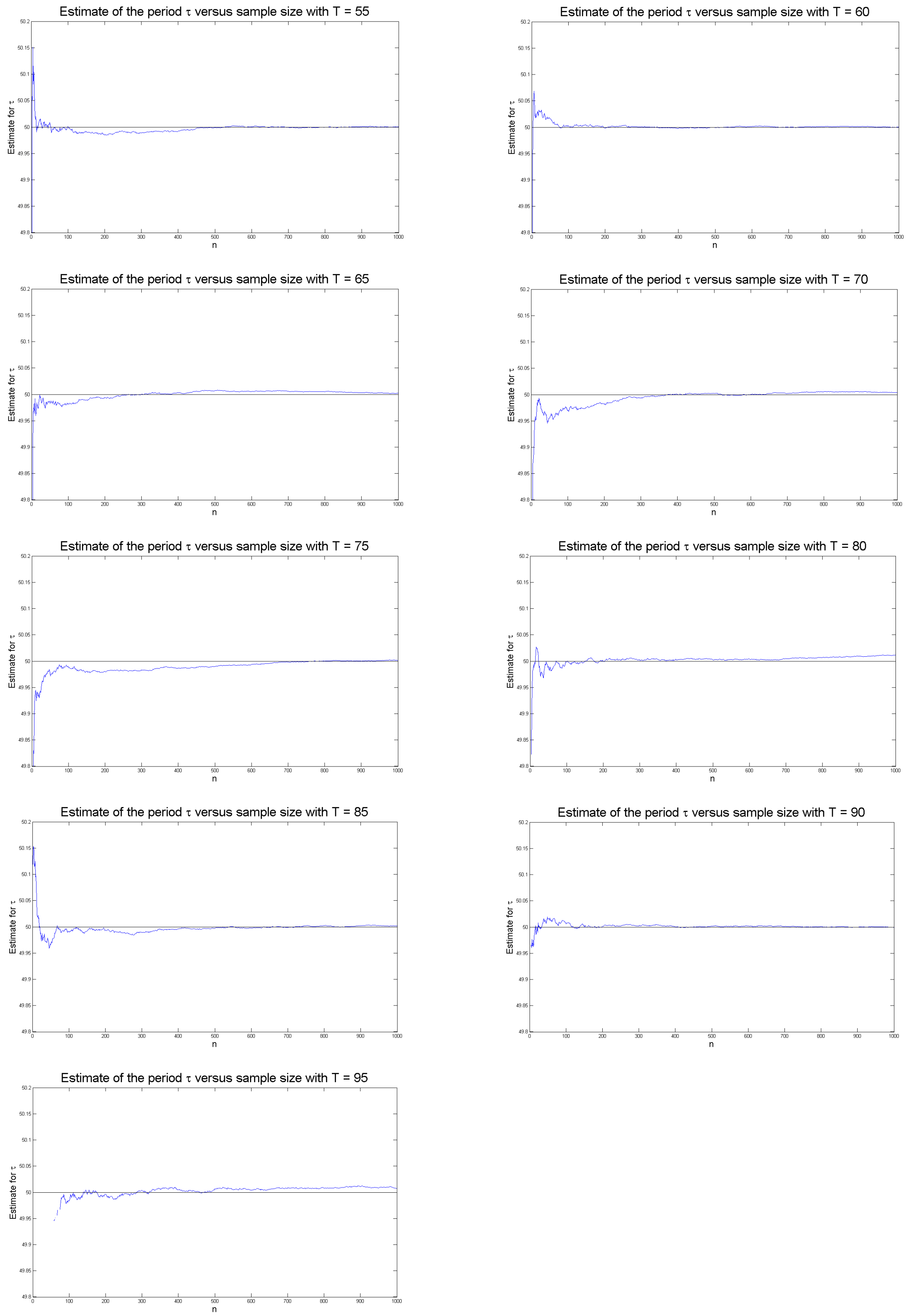


Fig. 6.7: Estimates of the period τ for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom) as a function of the sample size.

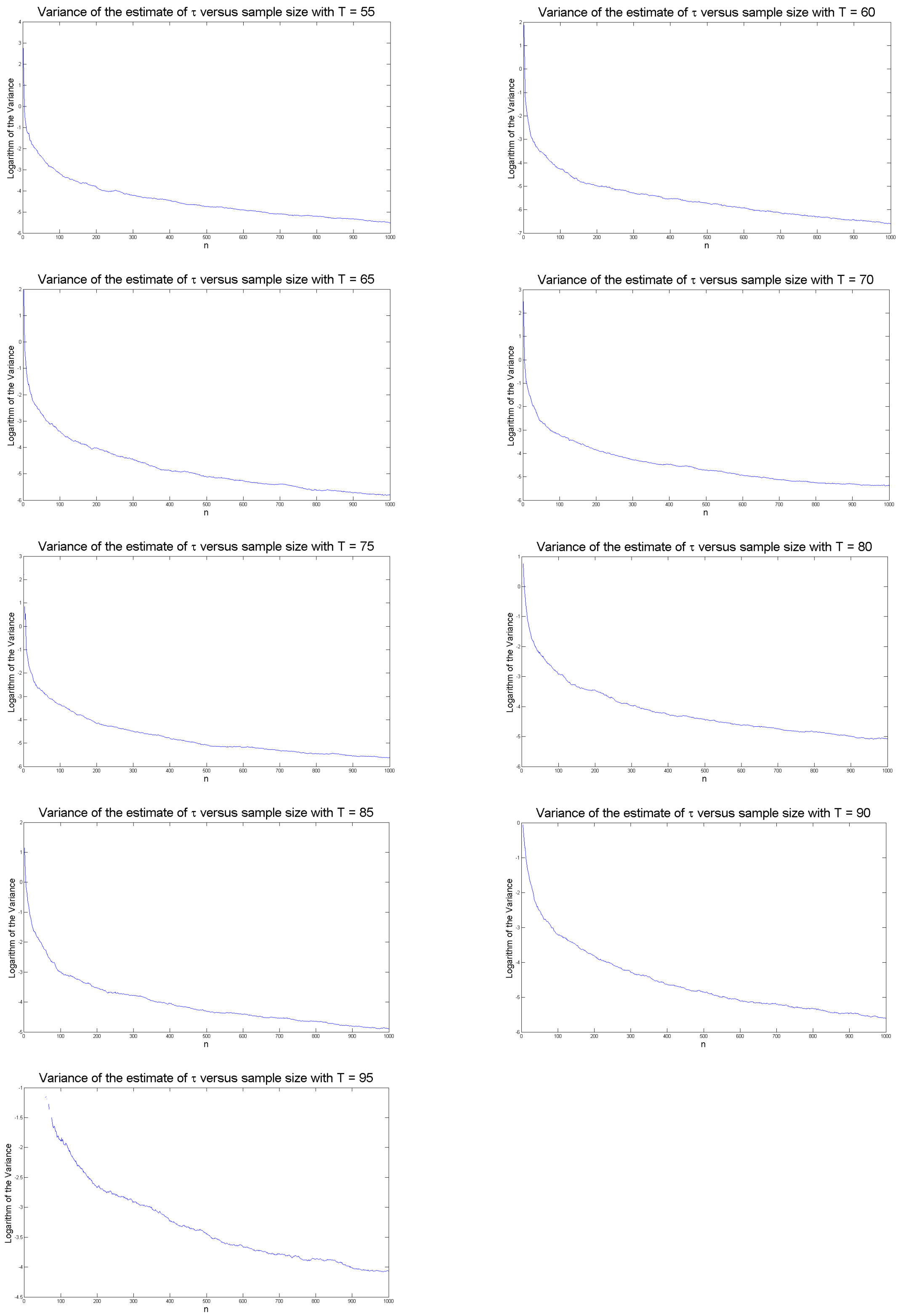


Fig. 6.8: Logarithm of the variance of the estimates of the period τ for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom) as a function of the sample size.

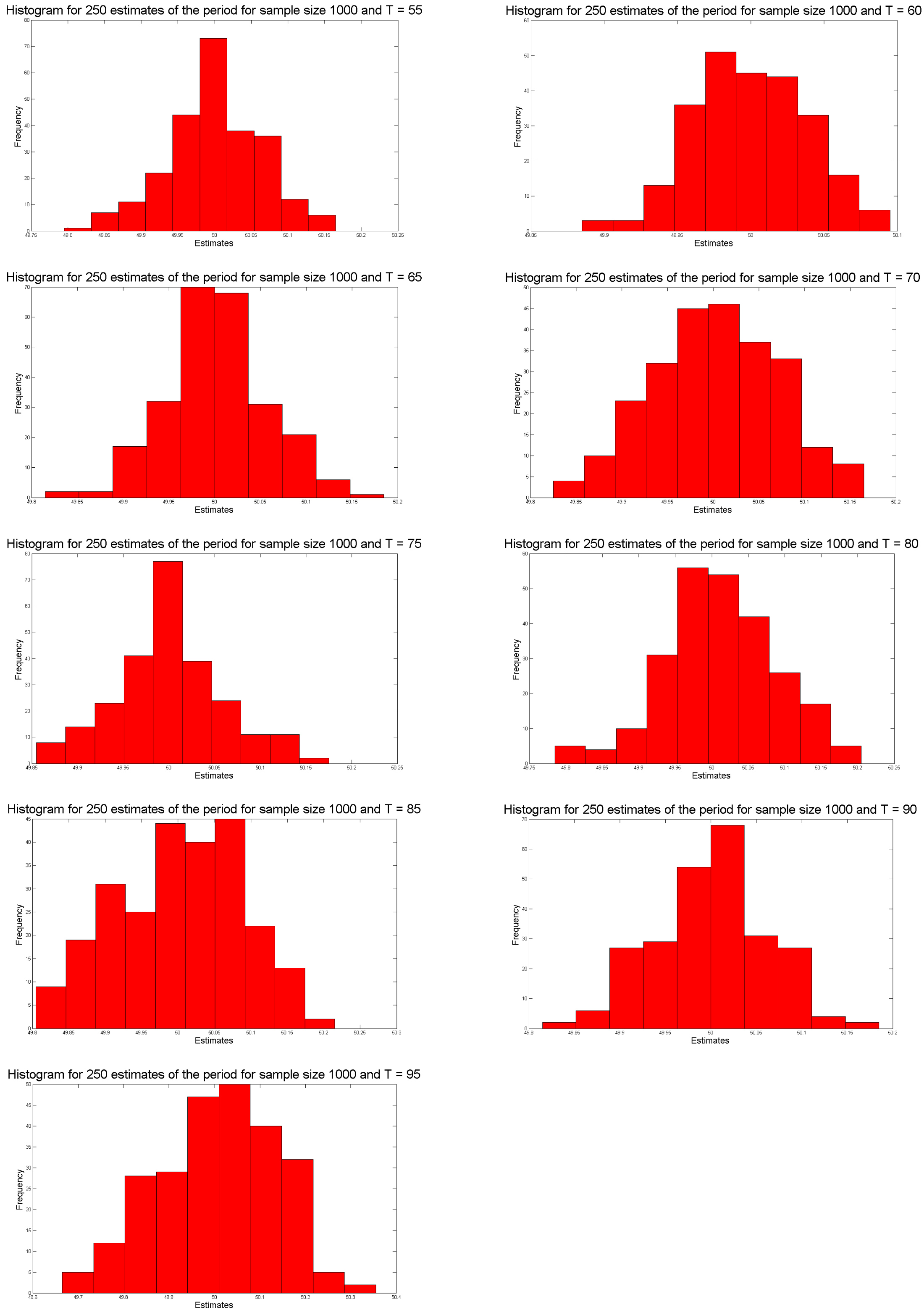


Fig. 6.9: Histogram for the final estimates for the period τ (sample size 1000) for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

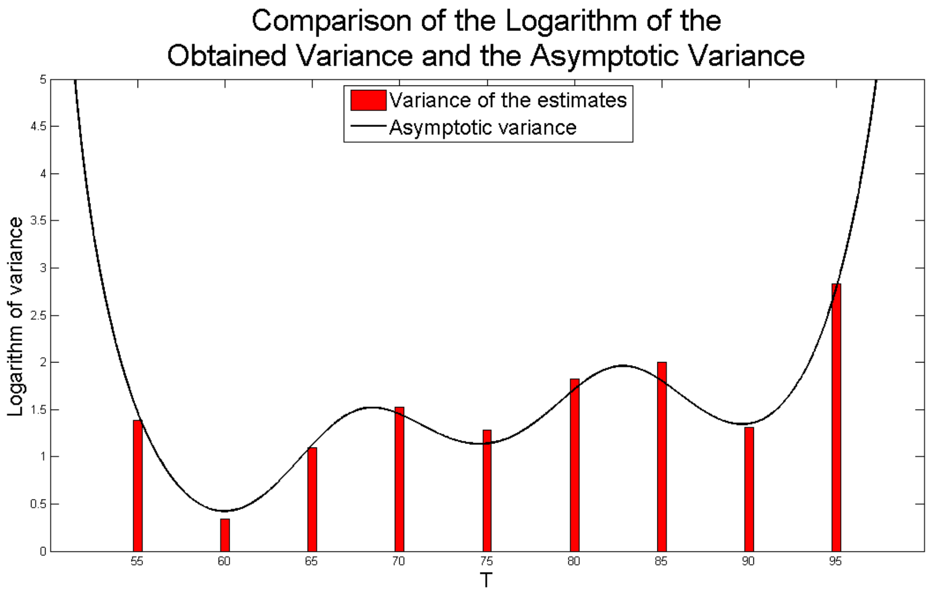


Fig. 6.10: Logarithm of the obtained variances the repeated estimates for several values of T and respective logarithms of the asymptotic variance obtained using (4.20).

6.3.2 Second version of the estimator

We will now take a closer look at the behaviour our estimator in its second version can present, by repeating the estimation of the period, with different data, 250 times and average the results, always considering a sample size of 1000 for each individual estimation. We will be, as in the previous section, focussing on $T = 55, 60, 65, 70, 75, 80, 85, 90, 95$. We begin by presenting pointwise averages for the several criterion function that we obtain to have an idea of its asymptotic shape (Figure 6.11).

As with the first version of the estimator there is not much difference between the graphs in Figure 6.11 (with the mean of several estimations of Ψ) and the ones in Figure 6.2 (where we have just one estimate for Ψ). We note that for $T = 95$ we have a local minimum at $\theta_0 = 45$, as we would expect, but the function is so flat around that region that our implementation of the estimator has trouble finding the said local minimum. We therefore will exclude $T = 95$ from further analysis. We note as well that for $T = 75$ we have local minima other than just the one close to $\theta_0 = 25$. This sort of situations should, in practical situations, be handled in one of two ways: either we should choose another value for T or use a graph like Figure 6.3 to help us identify which of the local minima was the one we were interested in and then reduce the parameter set Θ. In this chapter we took the second approach when choosing the parameter set since we were mostly interested in seeing how the different choices for T affected the results.

The graphs in Figure 6.12 show the pointwise variance of our estimates of Ψ for the indicated values of T as a function of θ. We see that our normalisation factor seems to have broken the almost linear, increasing, trend we saw in the previous section: except for the first two graphs in Figure 6.12 it does not even seem reasonable to say that the variance increases with θ. The variance does seem to achieve its lower values close to the correct value for the parameter. This indicates – as always, at least for our choice for λ – that the actual variance, in a neighbourhood θ_0, is being underestimated by using the expression for the variance at θ_0 since we would expect the variance to be roughly constant at least close to θ_0. This was of course expected since the expression for the variance of Ψ was only established at $\theta = \theta_0$ so we already knew that we would either be underestimating or overestimating the variance for other values of the parameter: the graphs suggest that in our present case we are practically always underestimated the variance at these points. Note also that the variance for the correct value of the parameter is close to $1/1000$ suggesting that the variance in indeed being properly normalised.

Next, we look at the evolution of the average of the estimates as a function of the sample size (Figure 6.13). We notice in Figure 6.13 that even though the estimates seem to be converging it does look like there is some bias that results in the period being overestimated. Also the range of the average of the estimates is higher the one presented by the first version of the estimator.

We present now the logarithm of the variance of the estimates as a function of the sample size (Figure 6.14). Figure 6.14 sows us the variance of the estimates produced by the estimator is reduced as the sample size is increased regardless of or choice for T. With the exception of the sudden changes that some of the graphs present, t he way in which the variance progress does not seem to change much as we modify the value of T; the sudden drops in the variance are due to the implementation of the algorithm.

We also present for each value of T an histogram of the final estimates for the period (Figure 6.15). The histograms for the several estimates that were obtained do not seem to be much different from the ones we saw for the first version of the estimator in Figure 6.9. We also see like it the past section what looks to be some asymmetry in the graphs.

Table 6.4 summarises the results: it contains the p-values for the Kolmogorov-Smirnov normality tests as well as the obtained average estimates, variance and kurtosis.

We see than in no case can we reject (at a level of, say, 5%) the hypothesis that the estimates produced by a sample of size 1000 are normally distributed. We also see that the fact that we no longer have a common distribution among the elements of our sample takes away some of the precision of the estimates and increases the variance of the results. We also notice that the average estimate for the period always exceeds the true value of the period.

T	Average Estimate	Relative Deviation	Variance	Kurtosis	p-Value
55	50.0580	0.1160%	0.0244	2.6110	0.7918
60	50.0615	0.1230%	0.0122	3.0188	0.2917
65	50.0623	0.1246%	0.0461	2.9581	0.6698
70	50.0492	0.0985%	0.0334	2.8378	0.9659
75	50.0540	0.1079%	0.0092	2.8851	0.6727
80	50.0585	0.1170%	0.0447	2.9487	0.9246
85	50.0710	0.1420%	0.0784	2.9626	0.5258
90	50.0282	0.0565%	0.0280	3.0089	0.2733

Tab. 6.4: Summary of the obtained results for the second version of the estimator.

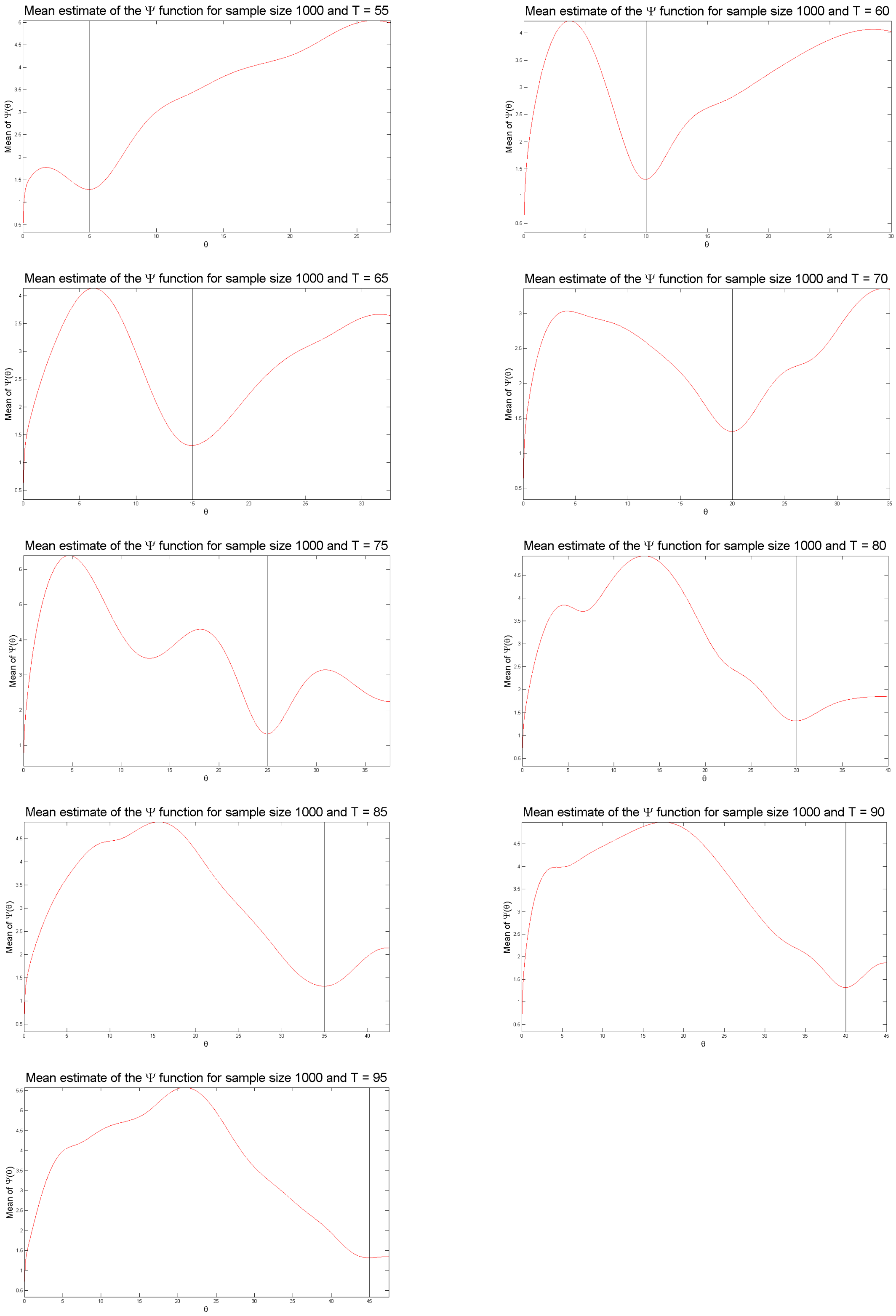


Fig. 6.11: Pointwise mean of the estimates of the Ψ function for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

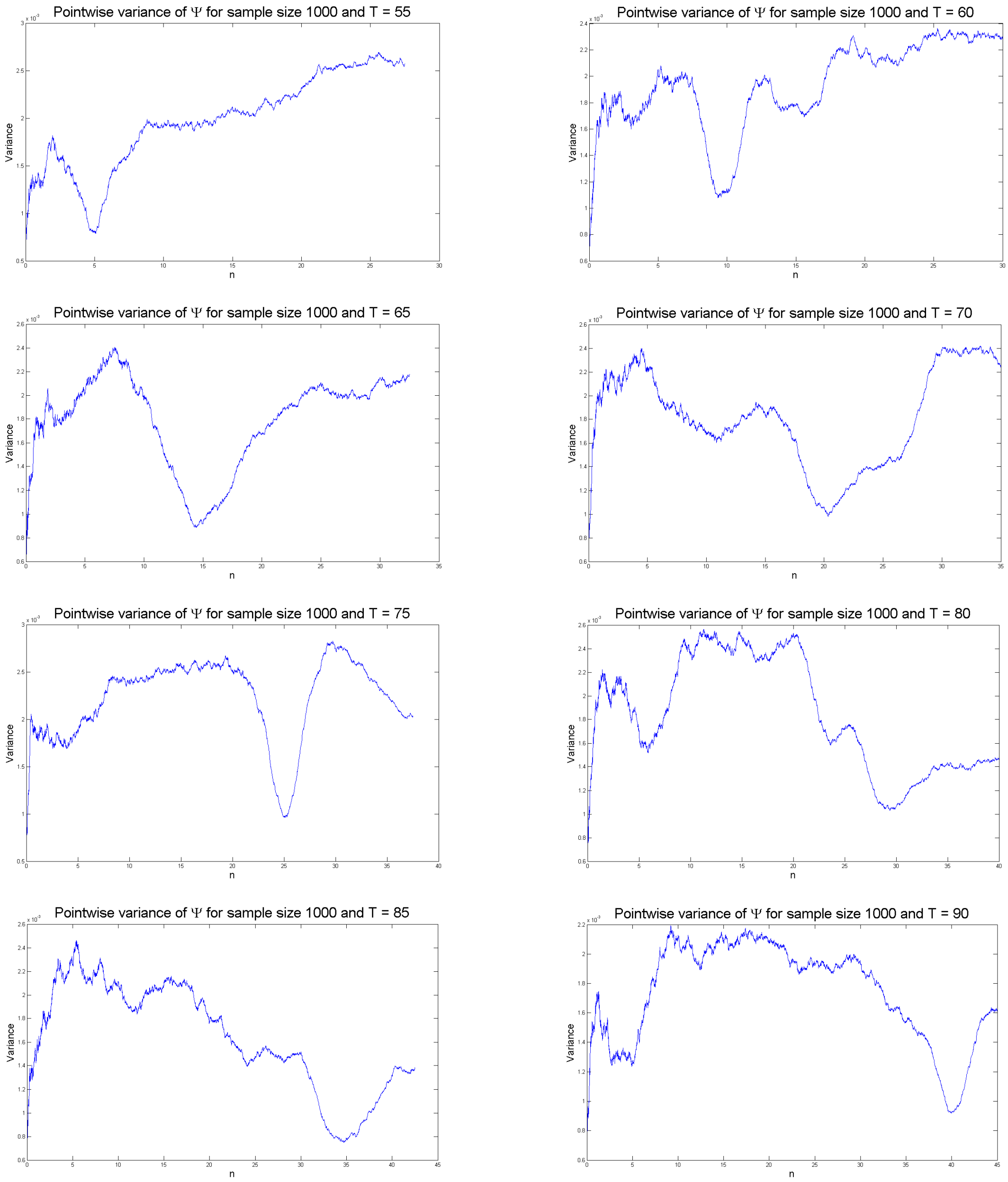


Fig. 6.12: Logarithm of the variance of the estimates of the Ψ function for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

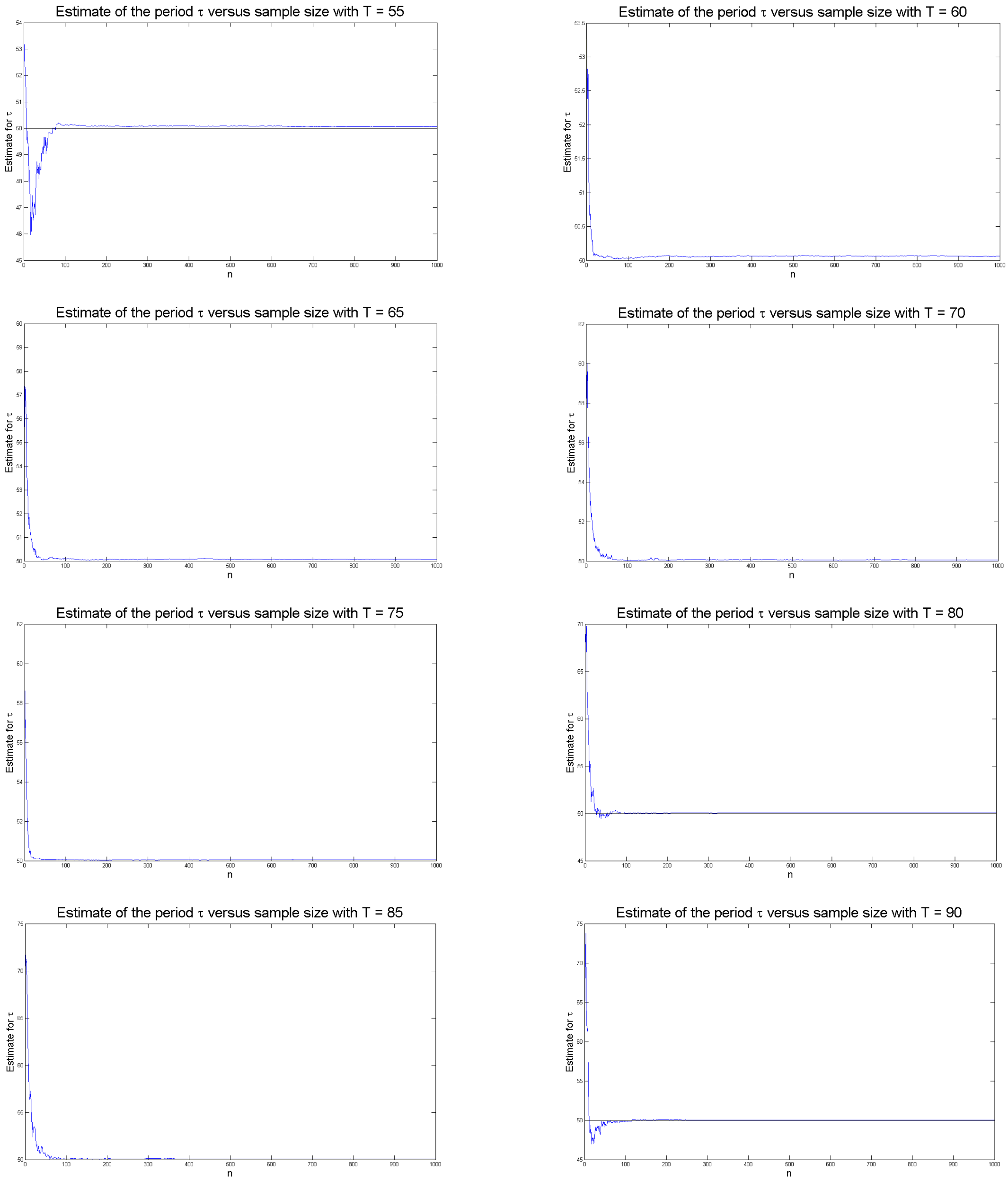


Fig. 6.13: Estimates of the period τ for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom) as a function of the sample size.

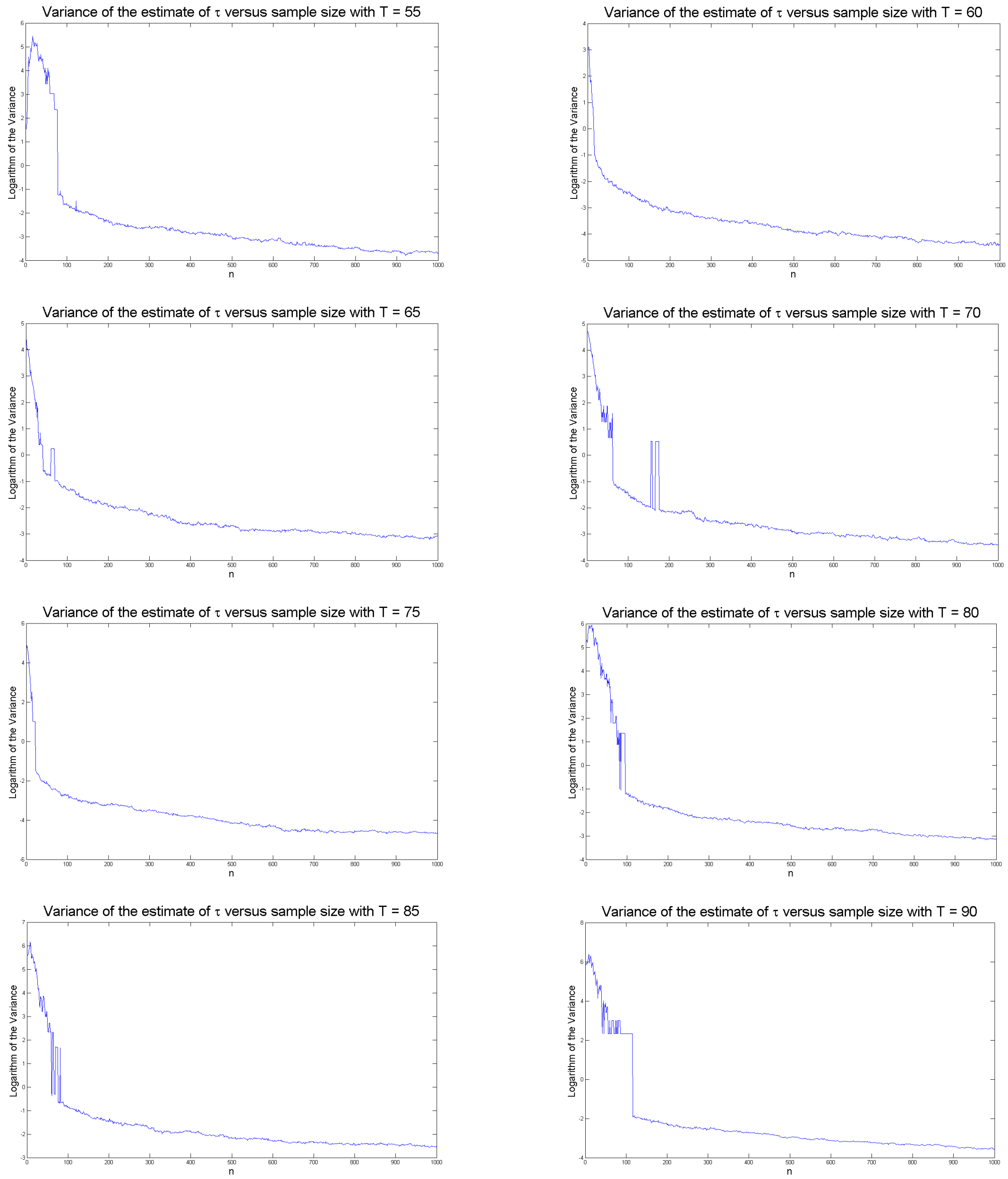


Fig. 6.14: Logarithm of the variance of the estimates of the period τ for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom) as a function of the sample size.

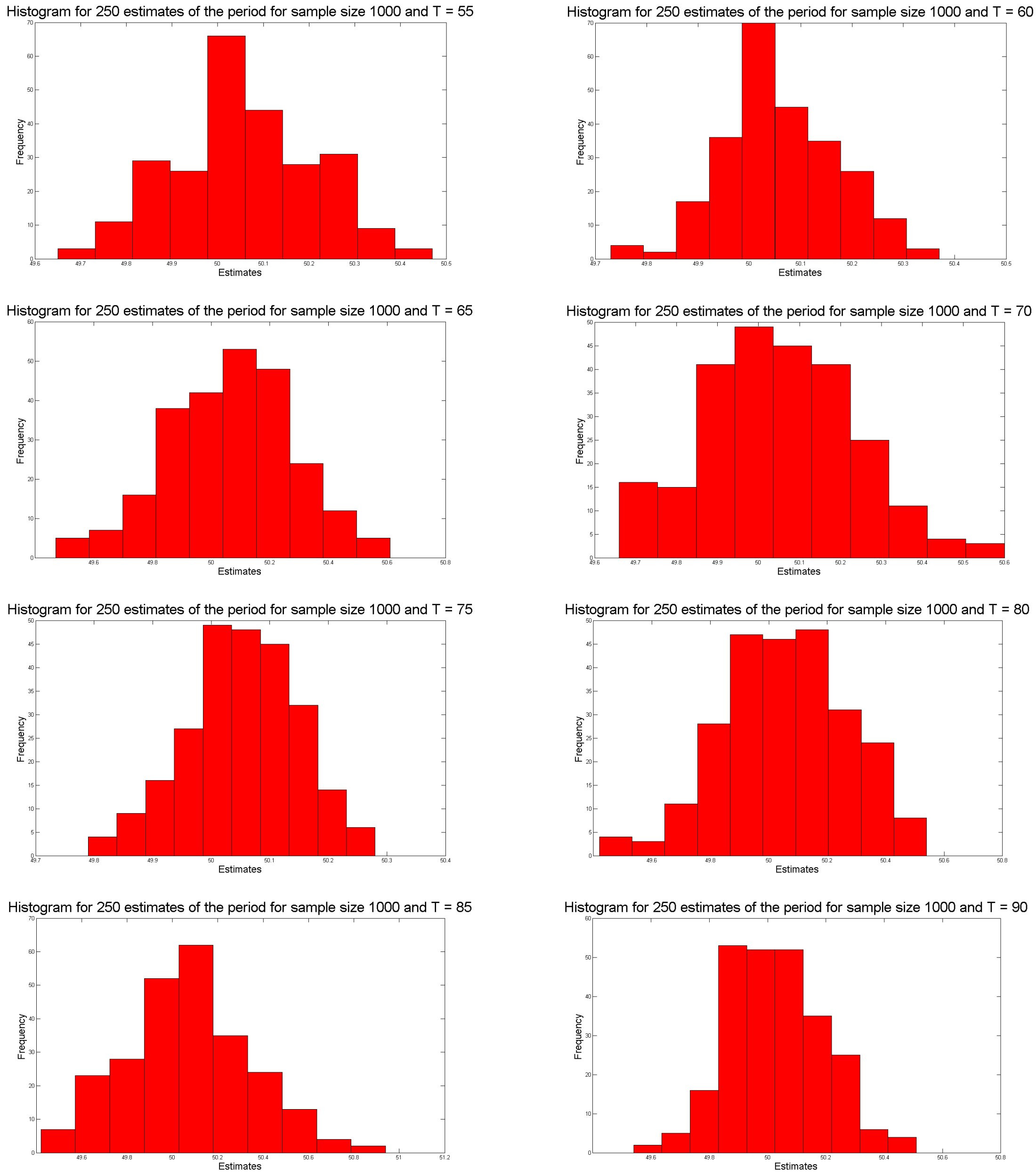


Fig. 6.15: Histogram for the final estimates for the period τ (sample size 1000) for T is 55, 60, 65, 70, 75, 80, 85, 90 and 95 (left to right and top to bottom).

Up until now we have been working with a fixed number of events per cycle. In the next section we will be modifying the average number of events per cycle. We will try see how this affects our results.

6.4 Effect of the Number of Events on the Estimation

For a fixed time interval, working with an intensity function whose integral over one period (read expected number of events per period) is large will give us more information to work with and should in principle allow us to get better estimates. There are of course many ways in which we could look at the influence that of the number of events per period has on the estimation procedure – say simply considering different intensity functions; since we are here considering a fixed intensity function, however, it seems more natural to just scale it as a form of controlling the number of events per period (50 time units).

In this section we will be looking at the influence that scaling the intensity function has on the estimates produced by our estimator. We will still be working with the same intensity function but it will be scared up and down. Until now, the intensity function had a maximum value of 10; in the following we will scale the intensity function to obtain intensities within our considered family with maximum intensities of $1.25, 2.5, 5, 7.5, 10, 12.5$ and 15 which correspond to an average of $25, 50, 100, 150, 200, 250$ and 300 events per period respectively.

6.4.1 First version of the estimator

We begin by considering the first version of our estimator. We will be, like in the previous two sections, repeating (100 times now) the estimation of the period but this time for each scaled intensity function. Figure 6.16 contains the obtained results in graphical form.

Figure 6.16 contains for each choice of the intensity function boxplots for the obtained estimates for increasingly larger sample sizes. In every case the range of the estimates that are produced decreases as we increase the sample size and the estimates are centred somewhere around 50. Perhaps more interesting is to compare the last boxplots in each of the graphs in the figure.We represent this in Figure 6.17.

Now we see (Figure 6.17) the effect that having more events per cycle has on the estimation: the range of values that the estimator can produces is reduced. This was already noted when looking at (4.20).

We further present (in Figure 6.18) the pointwise means for the final criterion function for each scaled intensity function. The functions in Figure 6.18 present the expected behaviour: the criterion functions are scaled by the same factor as the intensity function was. Also as expected (and desired) all the criterion functions intersect at $\theta_0 = 10$, the correct difference between our choice of $T = 60$ and the period $\tau = 50$ and therefore the zero of the limiting criterion function. We see also that in the vicinity of the zero the function is changing its values faster for the criterion function corresponding to the processes with higher values intensities. This indicates that the derivative of the criterion function at $\theta = T - \tau$ - which in this context is proportional to the inverse of the asymptotic variance of our estimator – increases as the intensity is scaled up: it actually increases in the same factor as already mentioned in Section 4.5. We now summarise the results that we obtained in this section in Table 6.5.

The results for the variance seem to fit those expected based on (4.20) so, they naturally decrease as the expected number of events per cycle increases; the relative deviations seem to present a decreasing trend although not monotonously so; in all cases we do not reject the hypothesis that the estimates obtained from samples of size 1000 is normally distributed (at a level of 5% for example).

While knowing the effect that scaling the intensity function has on the estimator is of low practical use (since the intensity function is fixed), it is nonetheless interesting to see, on one hand, what this effect actually implies and, more importantly, that our estimator behaves in this

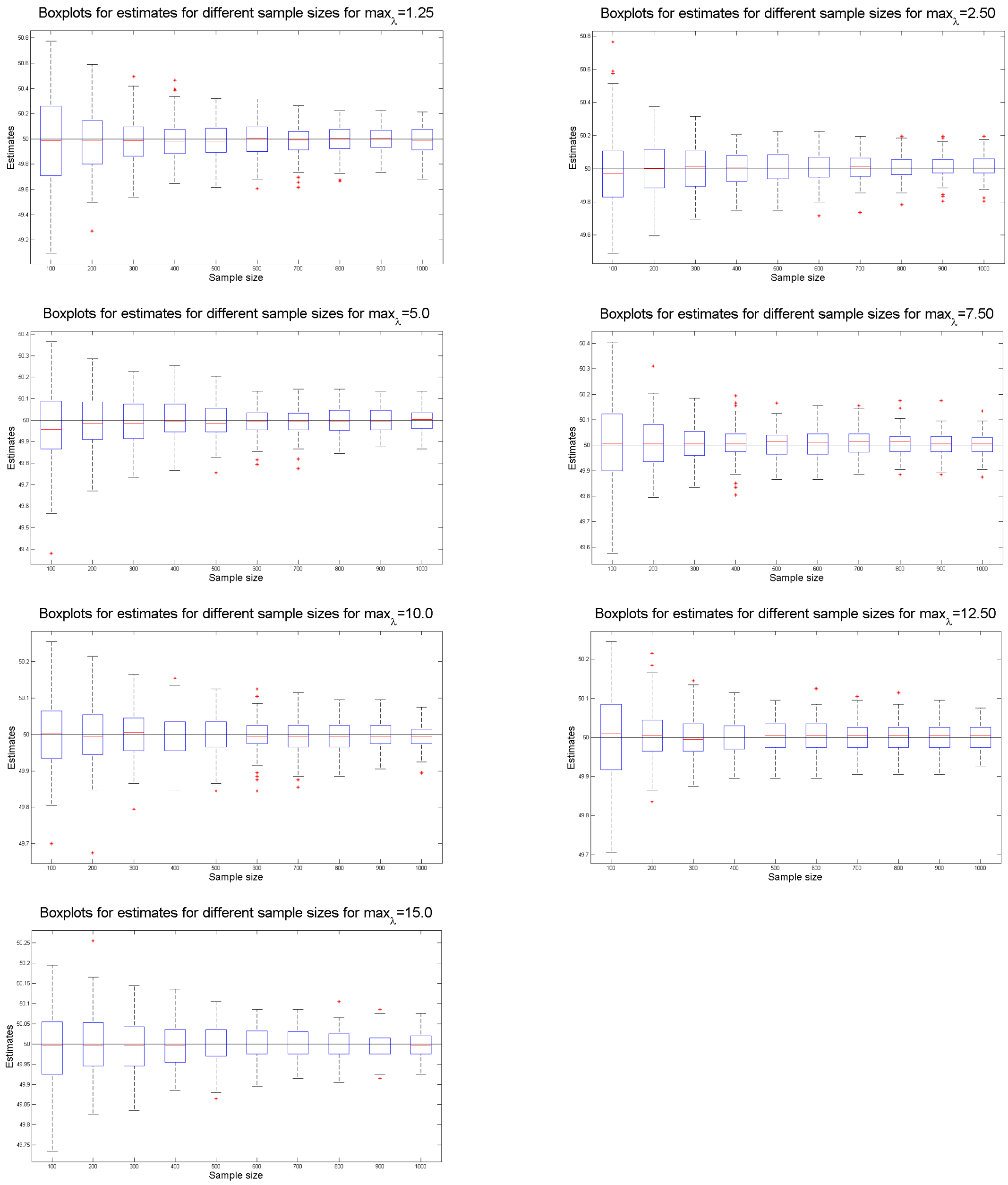


Fig. 6.16: Boxplots for the results of the repeated estimation of the period for increasingly larger samples for 25, 50, 100, 150, 200, 250 and 300 expected events per period (left to right and top to bottom).

situation in accordance to the theoretical model.

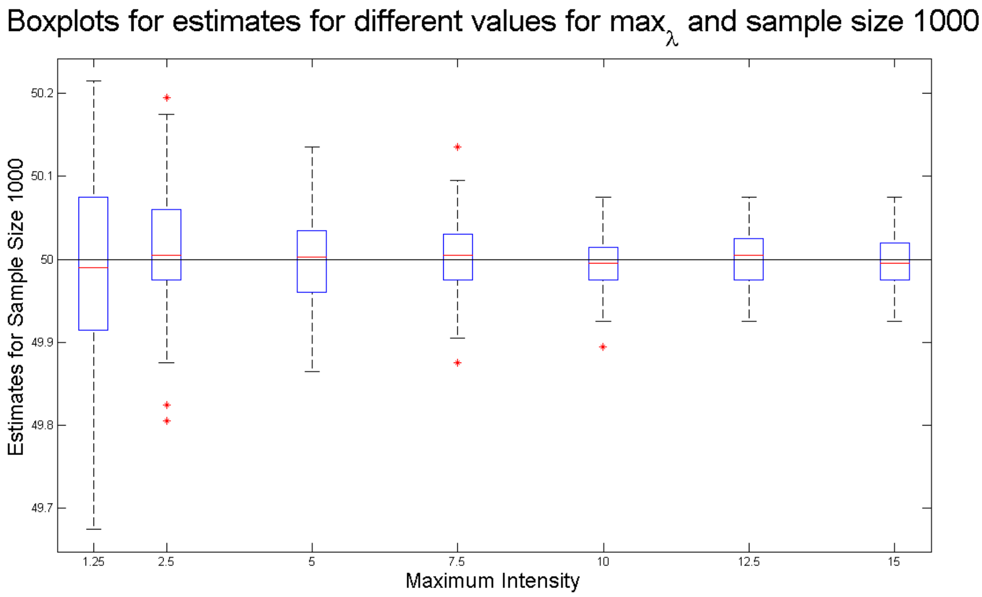


Fig. 6.17: Boxplots for the 100 estimates for the period obtained for repeated estimation with samples of size 1000 as a function of the maximum intensity.

Average Number of Events per Cycle	Average Estimate	Relative Deviation	Variance	Asymptotic Variance	Kurtosis	p-Value
25	49.9930	-0.0141%	0.0126	0.0122	2.6664	0.6919
50	50.0117	0.0234%	0.0052	0.0061	3.5324	0.1911
100	49.9986	-0.0027%	0.0036	0.0031	2.7428	0.5174
150	50.0025	0.0050%	0.0019	0.0020	3.3558	0.5622
200	49.9970	-0.0060%	0.0012	0.0015	2.9827	0.1858
250	49.9989	-0.0022%	0.0012	0.0012	2.4388	0.3905
300	49.9966	-0.0069%	0.0011	0.0010	2.5878	0.3523

Tab. 6.5: Summary of the obtained results for the first version of the estimator.

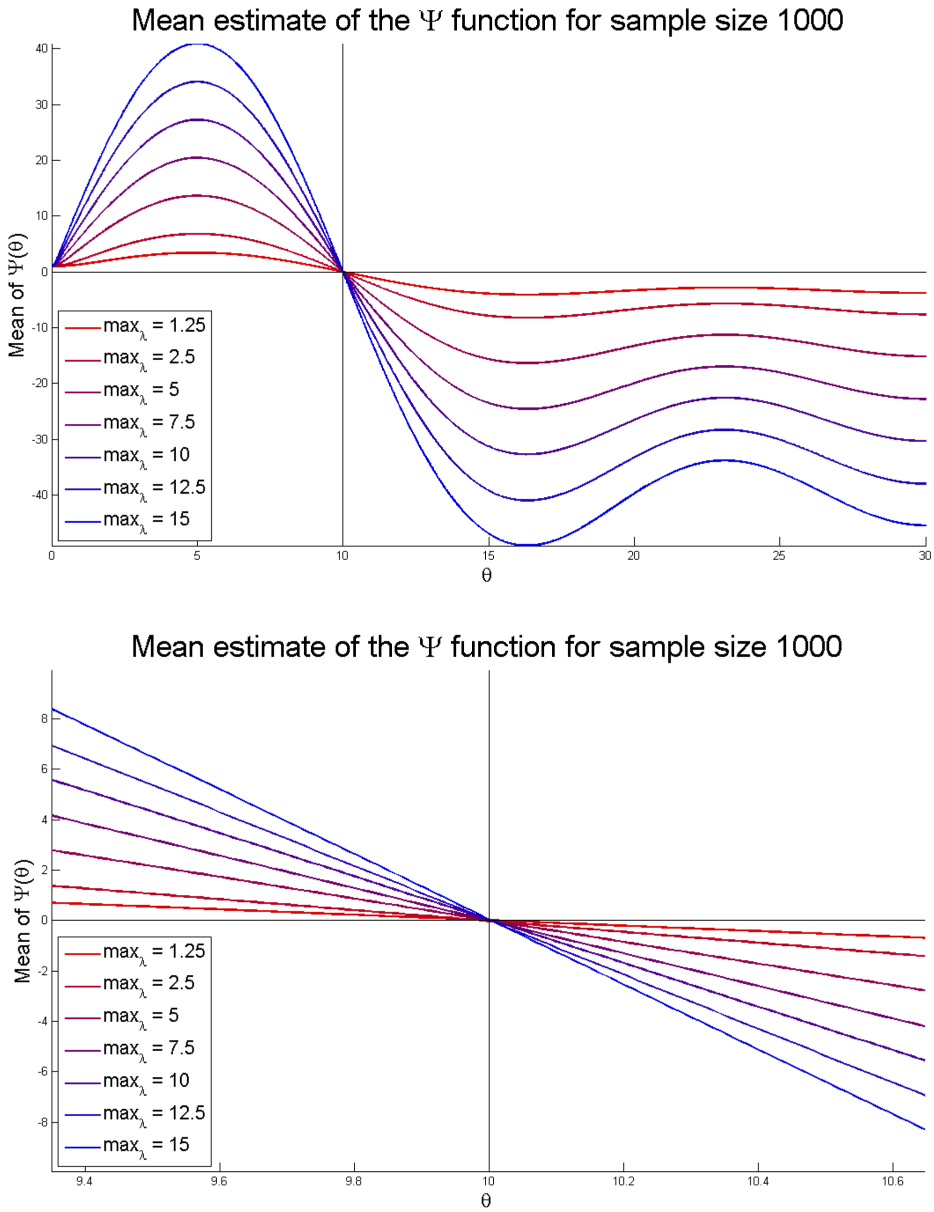


Fig. 6.18: Pointwise mean of the criterion function obtained from 100 repetitions of the estimation using samples of size 1000 for the several scaled intensity functions.

We will in the following section repeat the same tests for the second version of the estimator.

6.4.2 Second version of the estimator

We repeat in this section our experimentations regarding the average number of events per period in the intensity function Figure 6.19 contains the obtained graphs. They correspond to boxplots for the repeated estimates that are obtained for increasingly larger sample sizes for a fixed value for the maximum of the intensity function.

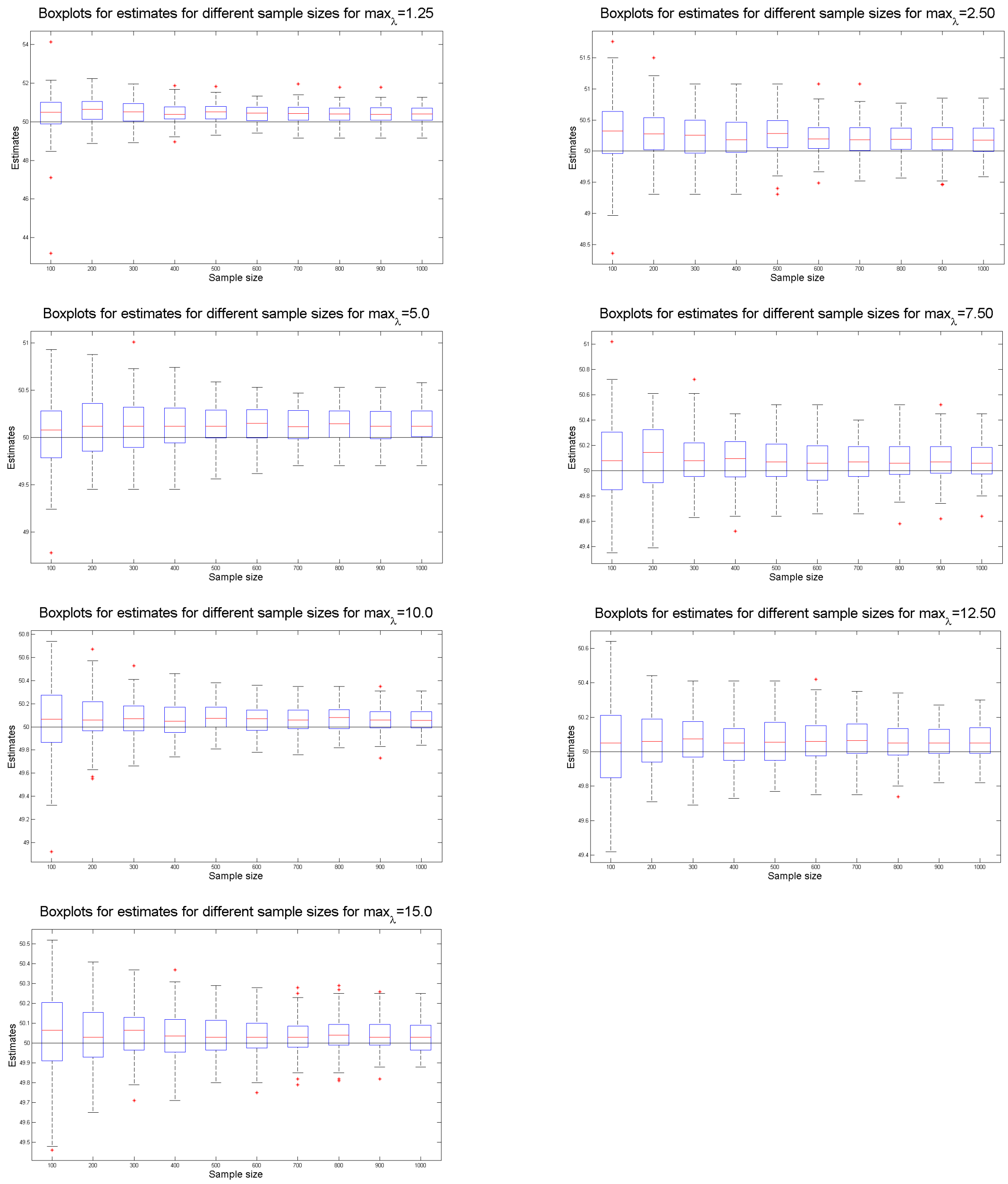


Fig. 6.19: Boxplots for the results of the repeated estimation of the period for increasingly larger samples for 25, 50, 100, 150, 200, 250 and 300 expected events per period (left to right and top to bottom).

We also gather the last boxplots in each of the graphs in Figure 6.19 to better compare the effect that scaling the intensity function has on the estimates; The resulting graph is in Figure 6.20.

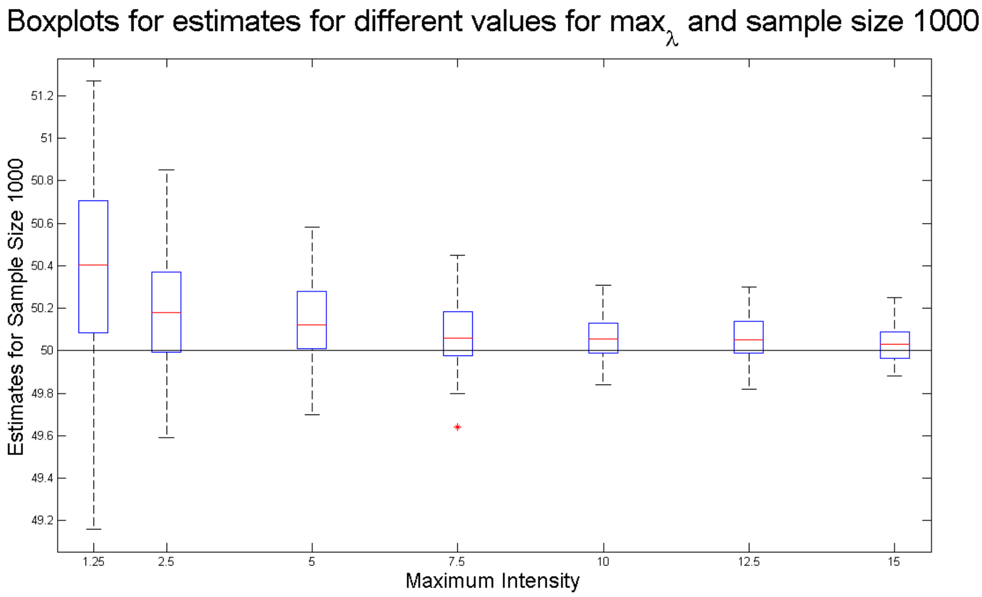
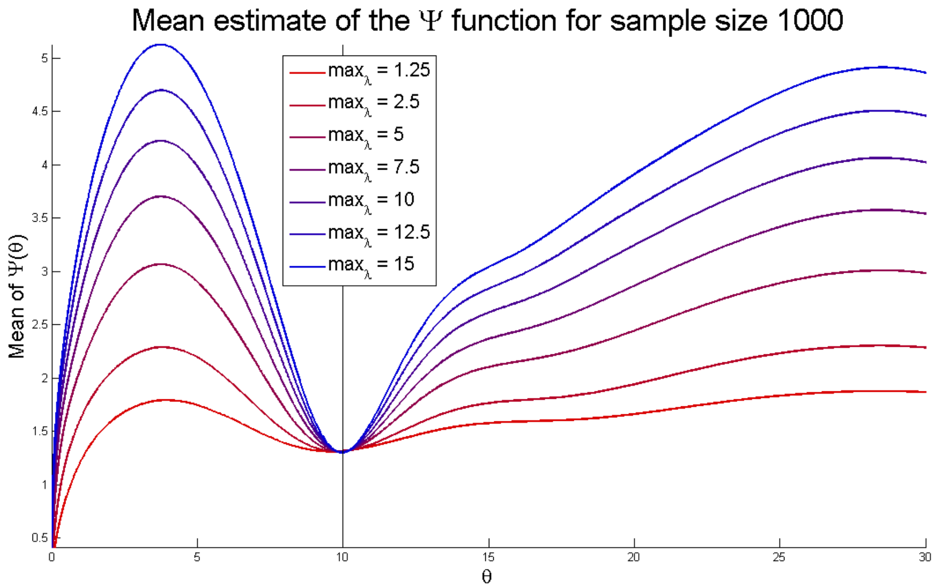


Fig. 6.20: Boxplots for the 100 estimates for the period obtained for repeated estimation with samples of size 1000 as a function of the maximum intensity.

Figure 6.20 shows that we seem to overestimating the period but that this tendency decreases as we increase the number of vents per period.

We also present (in Figure 6.21) the pointwise means for the final criterion function for each scaled intensity function. It is quite interesting to see that even though scaling the intensity function results in the corresponding criterion functions being scaled as well, the point where the local minimum for the criterion function is attained does not seem to change much both in terms of the value that is attained and the value of the parameter for which the local minima are reached. This is not however true if the normalising factor corresponding to the variance of the estimator is left out. Figure 6.22 shows the results. We see that not only do the points where the minimum are attained seem to be shifting to the left as we scale the intensity function down, and the corresponding values for the minimum are getting smaller, but also the criterion function is getting flatter making it more difficult to identify the local minimum. This, put together with the fact the denominator in the definition of the criterion function (5.3) only corresponds to the correct normalisation for $\theta = T - \tau$ might explain why the results produced by the estimator seem to be a bit off in terms of mean.

We now summarise the results that we obtained in this section in Table 6.6.



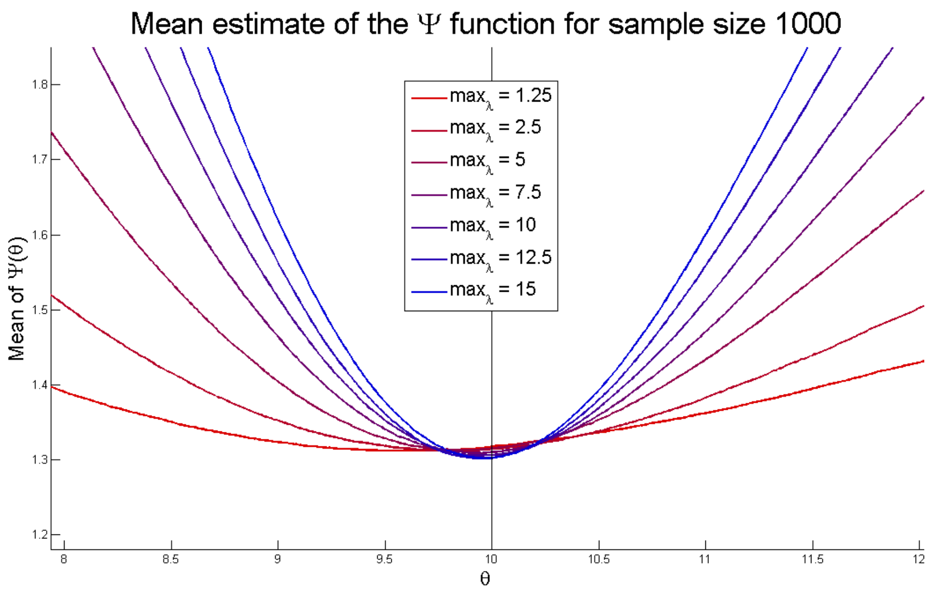


Fig. 6.21: Pointwise mean of the criterion function obtained from 100 repetitions of the estimation using samples of size 1000 for the several scaled intensity functions.

Average Number of Events per Cycle	Average Estimate	Relative Deviation	Variance	Kurtosis	p-Value
25	50.3911	0.7822%	0.1752	2.7400	0.9994
50	50.1871	0.3742%	0.0739	2.6730	0.7788
100	50.1286	0.2572%	0.0359	2.6027	0.9347
150	50.0680	0.1360%	0.0191	3.1387	0.8282
200	50.0564	0.1128%	0.0104	2.6225	0.5950
250	50.0615	0.1230%	0.0117	2.5045	0.8081
300	50.0341	0.0682%	0.0070	2.4508	0.3767

Tab. 6.6: Summary of the obtained results for the second version of the estimator.

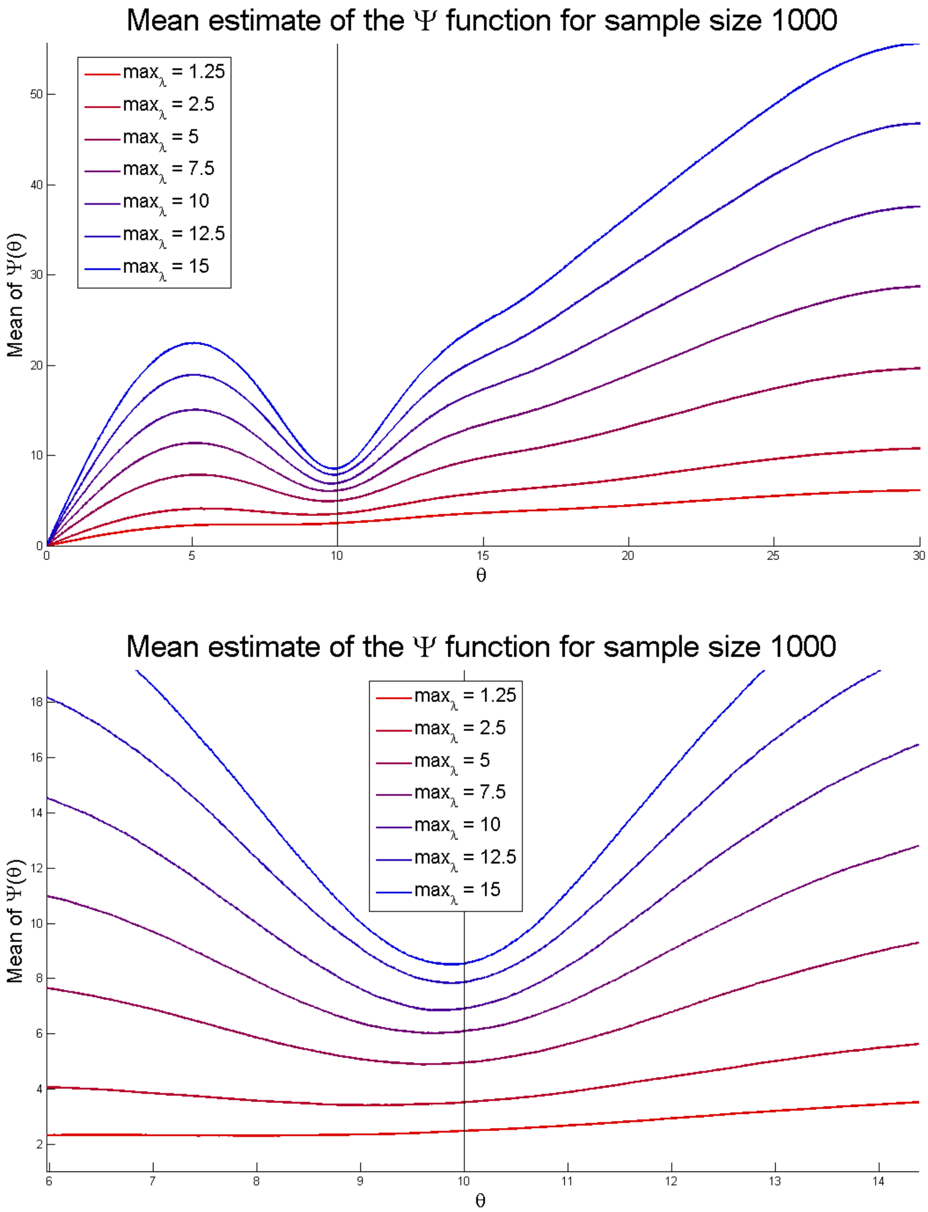


Fig. 6.22: Pointwise mean of the non normalised criterion function obtained from 10 repetitions of the estimation using samples of size 1000 for the several scaled intensity functions.

6.5 Comparison with known estimators

Since some experimental results were already known [15] for the particular case we treat here, we tried to reproduce some of the experimentation using our estimator so that it would present comparable results. We summarise this in Table 6.7. In our case each test was repeated 100 times. We present only results for the second version of the estimator since this is the version that deals with the type of data used in the tests.

Events per Cycle	Cycles	Total Events	Periodogram [11]	$\hat{\tau}_{n,max}$ [15]	$\hat{\tau}_{n,2}$ [15]	$\hat{\tau}_{n,3}$ [15]	Second Version	Average Deviation
25	40	1000	0%	6%	20%	24%	94%	3.2722%
50	40	2000	0%	38%	78%	78%	100%	1.8262%
50	20	1000	0%	4%	22%	32%	96%	2.6814%
50	10	500	0%	8%	12%	20%	83%	4.8574%
50	5	250	0%	8%	8%	10%	70%	7.8364%
100	10	1000	0%	30%	64%	68%	95%	3.0146%

Tab. 6.7: Summary of results produced by several different estimators (all counts of numbers of events should be read as expected counts).

According to the authors these tests were designed to see how the estimator performed under two situations: varying expected number of events per cycle for a fixed total expected number of events in the trajectory (first, third and last rows) and varying number of cycles observed for a fixed expected number of events per cycle (rows two through five).

Even though very small samples were used for these tests (compared to the ones used in the other sections of this chapter) our estimator seems to perform quite well not only outperforming all of the other estimators on all counts but maintaining a quite satisfactory rate of what the authors deemed "successful estimates". From the results it seems that our estimator is more affected (in a negative way) by a low number of events per cycle than by a sample consisting of a low number of cycles.

It is quite nice that the estimator is insensitive to how the events are distributed (more events per period for shorter trajectories or less events per period for longer trajectories) since we have no control over this: it doesn't seem good that the quality of the estimates varies so much for a fixed amount of information (1000 events per trajectory); its one thing for this to vary if we reduce the quantity of data but not if this remains constant.

In either case it seems quite impressive that our estimator can, with only five cycles being observed "successfully identify" the period 70% of the times versus 10% for the most effective estimator presented by the authors.

As for the results of the periodogram estimator, we will be referring to them in the next chapter (Chapter 7).

6.6 Other Intensity Functions

While an intensity function like (6.1) seems to be an interesting case study with its several peaks per cycle and the fact that the other estimators for the period seemed to present relatively bad results for it, it might be interesting to look at other intensity functions, namely estimates for Ψ for a few values of T and to look at the estimates for the period as a function of T. We use the same setting as in Section 6.2.1 and Section 6.2.2 changing only the intensity function being used to generate the data.

We recall that ρ indicates the maximum values of the intensity function and takes the value 10; τ will as always take the value 50.

6.6.1 Two step function

The first intensity function that we will consider here has an expected number of events per cycle of 375 and is given by (6.3).

$$\lambda(t) = \frac{\rho}{2} + \frac{\rho}{2}\left\lfloor 2\left(\frac{t}{\tau} - \left\lfloor \frac{t}{\tau} \right\rfloor\right)\right\rfloor \tag{6.3}$$

So we begin with a function that alternates between two levels – arguably the closest we can get to a homogeneous case. Since our estimator works by comparing estimated areas over several pairs of intervals, this case should pose serious identifiability issues since the function does not change much. In fact the pair of time intervals that we are comparing for each θ both have the same length, if in both of those integrals the function is at the same level, then the areas coincide; one would then expect there to be a myriad of local minima in Ψ_n. This does not happen as we can see in Figure 6.23.

We can see that the zeros still change as a function of T in an appropriate fashion. We notice however two anomalies: we have a spike at $T = 75$ an some instability for $T > 105$.

The spike at $T = 75$ is quite simple to understand; if we look at the intensity function in intervals of the type $[75(k-1), 75k]$ like the ones we are working with, we see that the intensity function is symmetrical within these interval and as such the integrals we are comparing should always coincide for all θ; this also explains the criterion function we observed for $T = 75$. This is of course a coincidence and would not happen for any value for T somewhat smaller or larger. The same peak can be seen at $T = 125$ for the same reason.

As for the instability for $T > 105$ it is due to regions of the criterion function like the one close to $\theta = 10$ that our implementation confuses with a local minima.

The other criterion functions we present behave as expected. We note however that Ψ_{1000} for $T = 90$ is quite flat close to $\theta = 40$; this causes our algorithm not to find any zeros which accounts for the absence of estimates close to the multiples of the period.

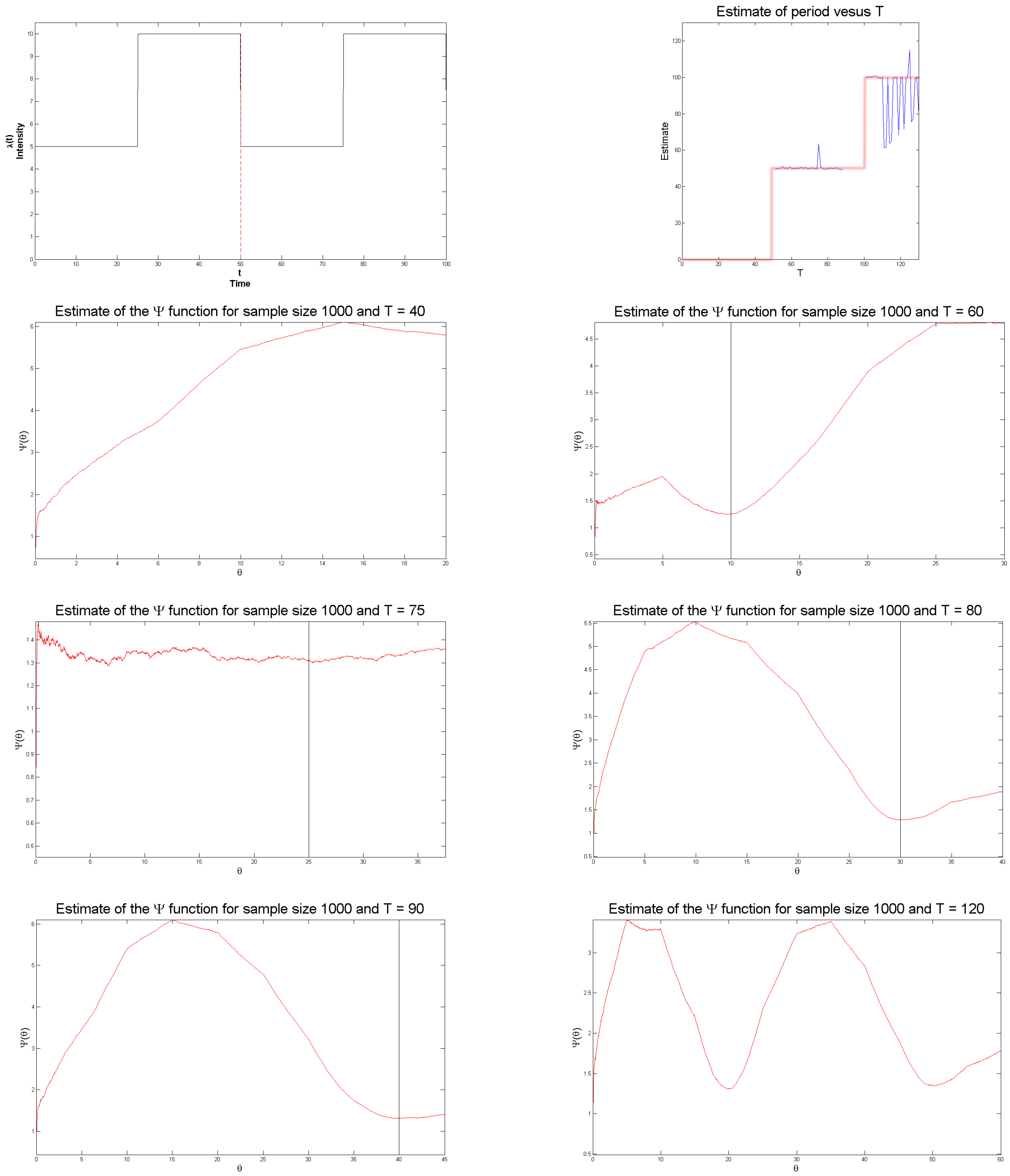


Fig. 6.23: Graphs for a two step function: graphical representation (top left); estimate for τ as a function of T (top right); Ψ_{1000} for several choices of T (indicated in the graphs' titles).

6.6.2 Three step function

We now look at the following intensity function that has an expected number of events per cycle of 312.5. We now have the intensity function taking three values.

$$\lambda(t) = \left(\frac{\rho}{2} + \frac{\rho}{2}\left\lfloor 2\left(\frac{t}{\tau} - \left\lfloor\frac{t}{\tau}\right\rfloor\right)\right\rfloor + \frac{\rho}{2}\left\lfloor 3\left(\frac{t}{\tau} - \left\lfloor\frac{t}{\tau}\right\rfloor\right)\right\rfloor\right)/2 \tag{6.4}$$

Figure 6.24 presents a graph of the intensity function and the results of the estimation.

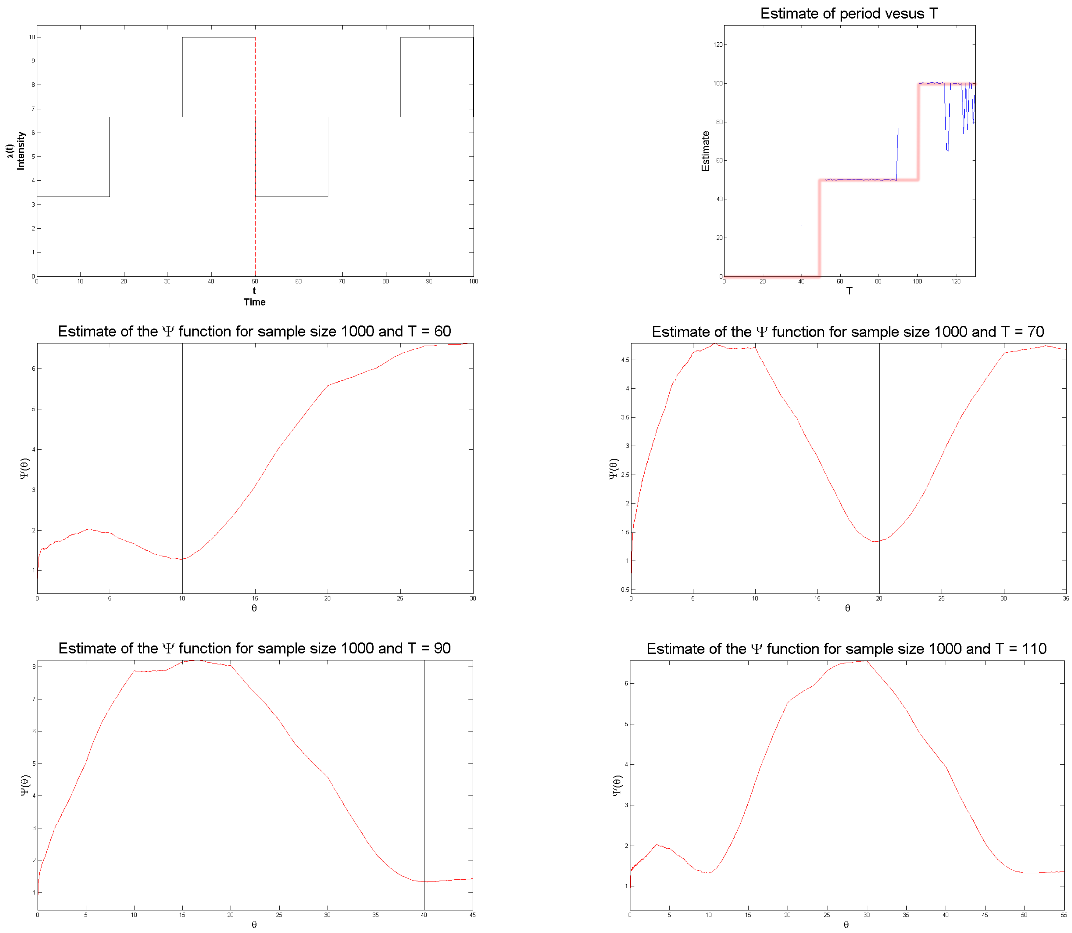


Fig. 6.24: Graphs for a three step function: graphical representation (top left); estimate for τ as a function of T (top right); Ψ_{1000} for several choices of T (indicated in the graphs' titles).

The introduction of a third step prevents the isolated peak that we saw in the previous sections to occur. Again, for some values of T larger than 100 we get some instability however, the fact that these peaks fail at holding the function at a constant level they do not make the identification of the period difficult.

6.6.3 Sawtooth

The third inspected shape for the intensity function is given by (6.5) and has an expected number of events per cycle of 250.

$$\lambda(t) = \rho\frac{t}{\tau} - \rho\left\lfloor\frac{t}{\tau}\right\rfloor \tag{6.5}$$

In Figure 6.25 we present a graph of the intensity function and the results of the estimation.

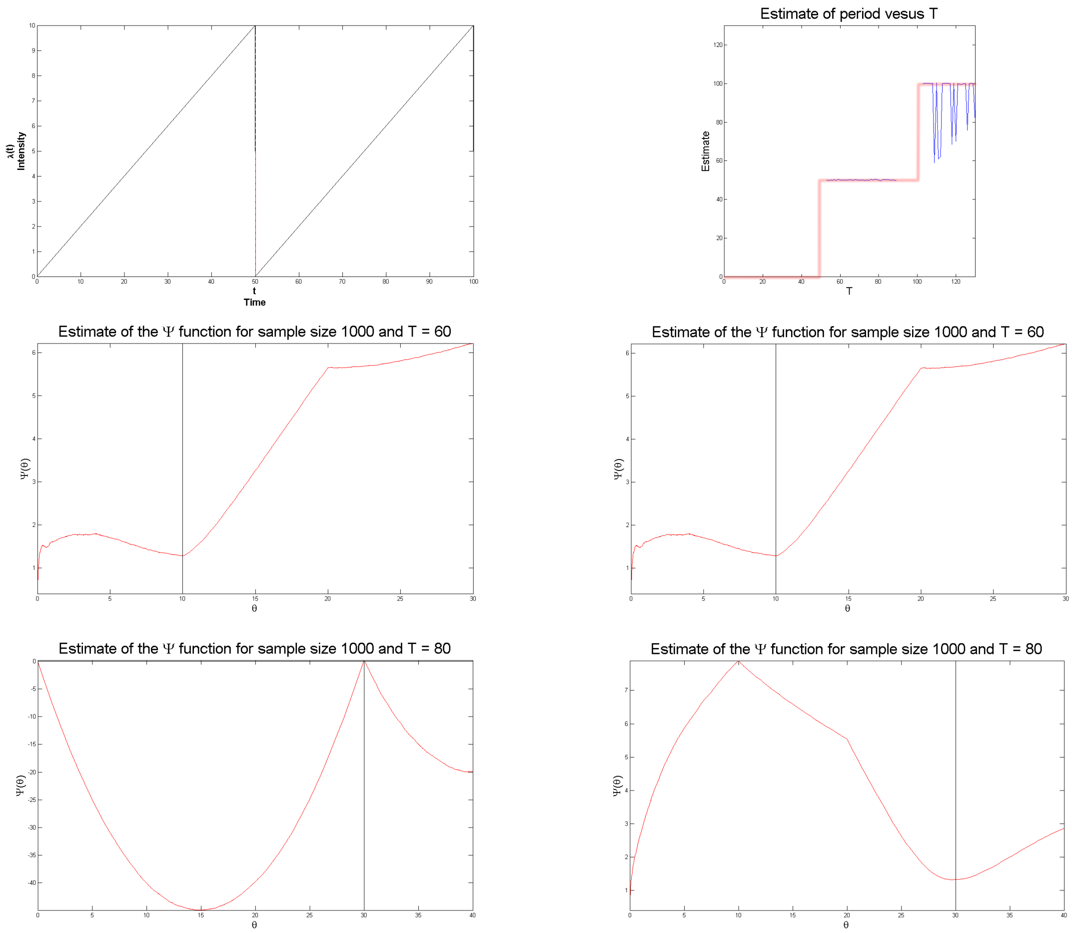


Fig. 6.25: Graphs for a sawtooth function: graphical representation (top left); estimate for τ as a function of T (top right); Ψ_{1000} for several choices of T (indicated in the graphs' titles). The graph on the bottom right corresponds to the first version of the estimator.

We follow the same pattern as before with some instability being shown for the step corresponding to twice the period. While understanding why, for low values of T (namely between the period and twice the period), the areas being considered (or estimates thereof) are close, when we start taking larger values of T it becomes much more difficult to understand when the areas will match.

The graphs in the last row of Figure 6.25 correspond to the criterion function for the first and second version of the estimator (to the left and to the right respectively). We present also the criterion function for the first version so show an example of how this function can have a zero without actually changing sign.

6.6.4 Sinus function

We now work with the sinus-type function (6.6). It has an expected number of events per cycle of 250.

$$\lambda(t) = \frac{\rho}{2} \sin\left(\frac{2\pi t}{\tau}\right) \tag{6.6}$$

In Figure 6.25 we present a graph of the intensity function and the results of the estimation.

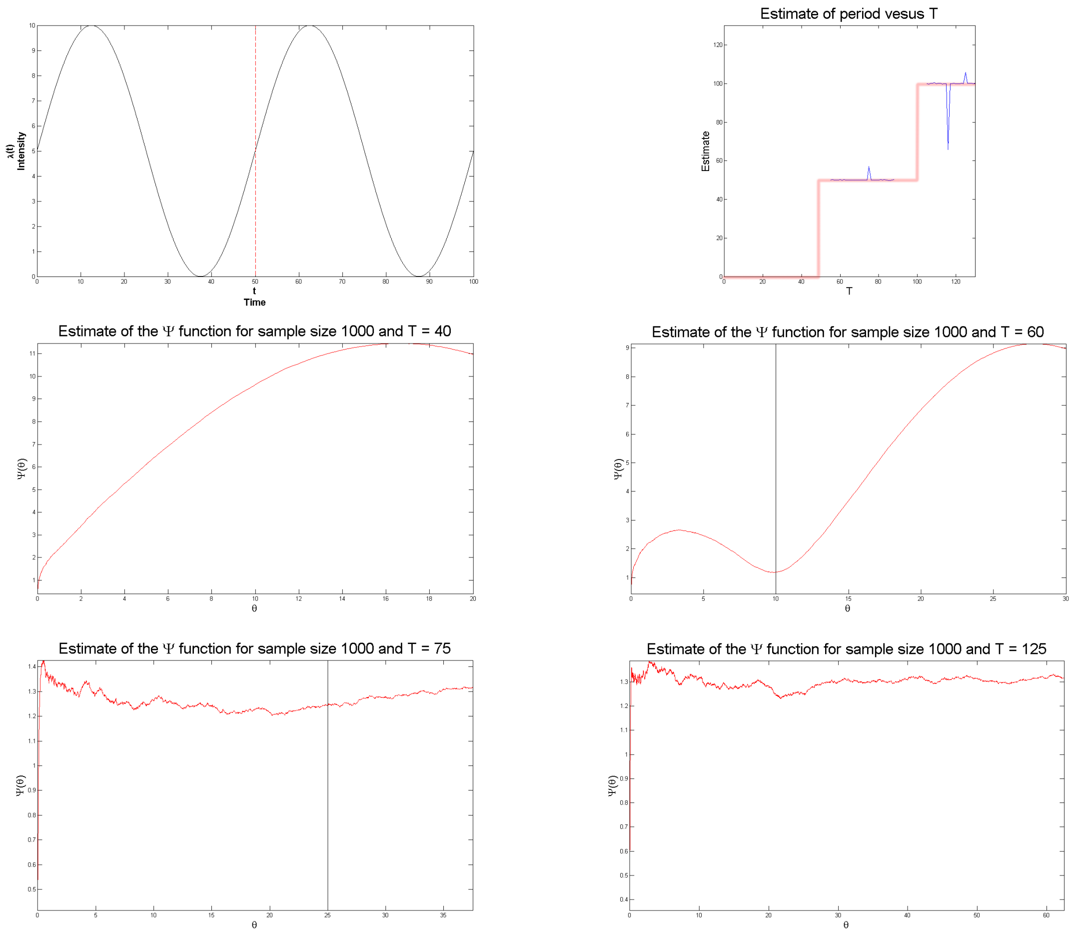


Fig. 6.26: Graphs for a sinus function: graphical representation (top left); estimate for τ as a function of T (top right); Ψ_{1000} for several choices of T (indicated in the graphs' titles).

We again see a peak at $T = 75$ and $T = 125$ as in Section 6.6.1 again due to the mentioned symmetries. The other peak again comes from the algorithm miss identifying the local minima.

6.6.5 From τ-periodical to $\tau/2$-periodical

We will be looking here at the intensity function given by (4.6):

$$\lambda_{T_0}^{\alpha}(t) = \begin{cases} \sin\left(\frac{2\pi}{\tau}(t' - (T_0 - \alpha - \tau))\right) + 1 & \text{if } t' \in [T_0 - \alpha - \tau, T_0 - \tau] \\ \sin\left(\frac{2\pi}{\tau}t'\right) + 1 & \text{if } t' \notin [T_0 - \alpha - \tau, T_0 - \tau] \end{cases} \tag{6.7}$$

We remind that each cycle of the function is simply a sinus function (6.6) where we replace the values taken by it in $[T_0 - \alpha - \tau, T_0 - \tau]$ with $\lambda'(t - T_0 - \alpha - \tau)$ with λ' given by (6.6). More interesting in this case than looking at our estimates for Ψ is to do something analogous to what was seen in Section 4.1.2 and see if it is indeed possible to identify the period.

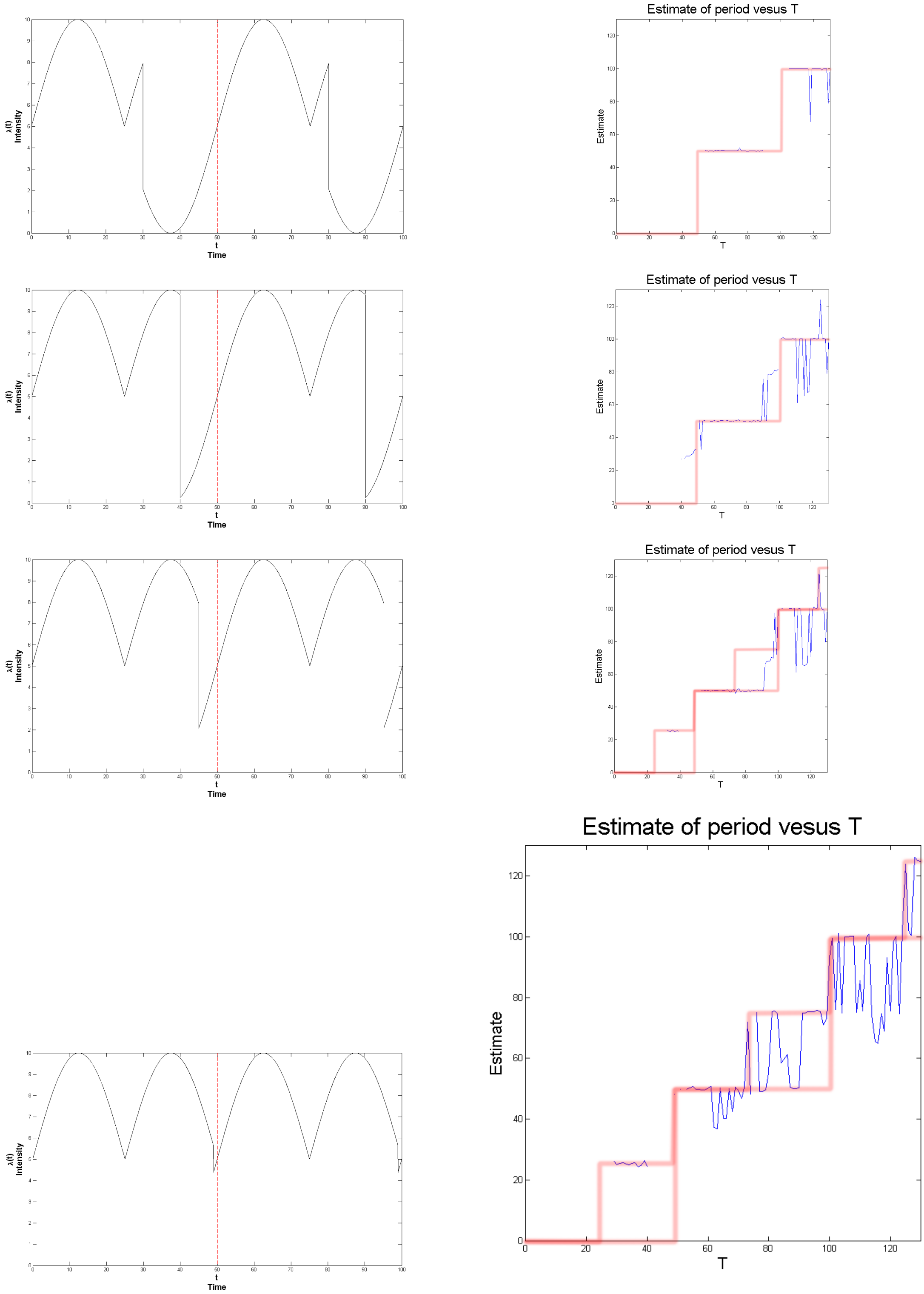


Fig. 6.27: Intensity function (on the left) and respective estimates for τ as a function of T (right side) for (α, T_0) in expression (6.7) taking the values $(5, 80)$, $(15, 90)$, $(20, 95)$ and $(24, 99)$ (per row).

This will be done by considering (α, T_0) taking the values $(5, 80)$, $(15, 90)$, $(20, 95)$ and $(24, 99)$ for which we have an expected number of events per cycle 265.198, 354.168, 393.957 and 408.527 respectively. Figure 6.27 contains the graphs for the intensity functions and respective graphs of the estimate for the period as a function of T.

We see that indeed the repetition of part of the intensity function does introduce an increasing amount of interference into the graphs as the repeated part extends; we can see that even when we are quite close the halving the period of the function as in the third row of Figure 6.27, the graph still seems quite treatable when we see other flat portions appearing; it is only for the last row that there could be indeed some confusion.

We further compare, however, in Figure 6.28, the graphs corresponding to the estimates as a function of T for (α, T_0) taking the values $(25, 99)$ making the intensity function almost $\tau/2$-periodical and $(25, 100)$ making it indeed $\tau/2$-periodical. We see that there are virtually no differences between the two graphs and that both a τ-staircase function and a $\tau/2$-staircase function seem to fit to the graph; then again there is not much difference to be seen between the two intensity functions so either conclusion would be reasonable.

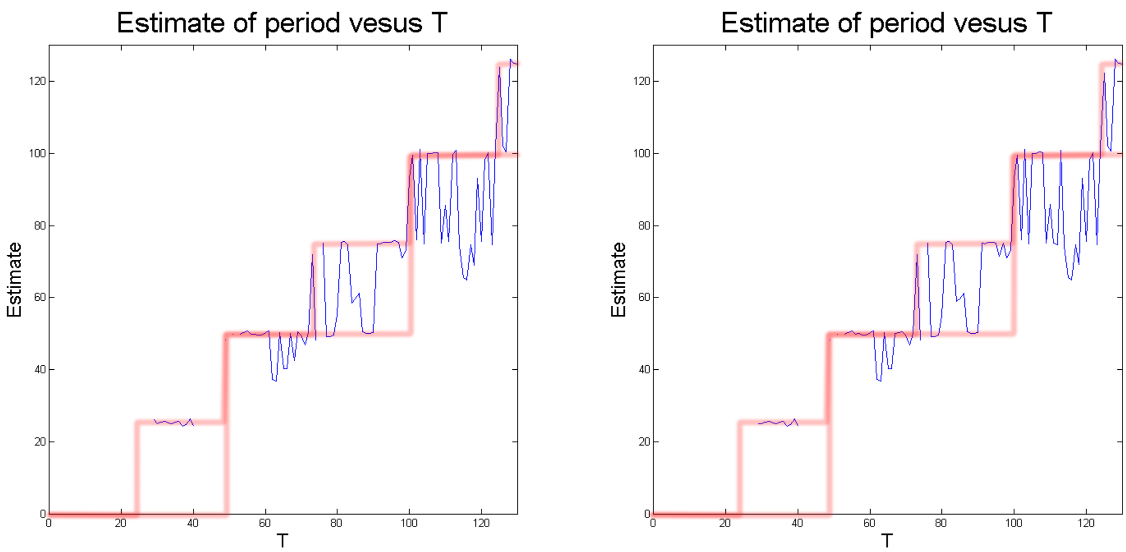


Fig. 6.28: Estimates for τ as a function of T for (α, T_0) in expression (6.7) taking the values $(25, 99)$ and $(25, 100)$ (left and right respectively).

There is also something else to notice. Given the discussion of Section 4.1.2 one would expect the interference of the repetition of the intensity function to extend more. This does not happen however for a simple reason; the reasoning that we had in the mentioned section and that guided us in concluding the expected shape for the graphs such as those in Figure 6.28; such a reasoning could not be repeated in this situation since due to the fact that we are partitioning a single trajectory we will, even for fixed T and θ be comparing different pairs of regions of the intensity function.

7. CONCLUSIONS

In this chapter we compile some conclusions and comments regarding the problem that was treated, previous estimators, our estimators, the implementation and experimental results that we obtained and some notes on possible extensions on our method.

We begin with some notes on the parametric estimator. Vere-Jones's periodogram estimator works under a quite restrictive assumption on the shape of the intensity function; an assumption that fails with just the presence of intensity peaks with different heights, for example. This estimator then falls into the "all or nothing" category either working quite well for a specific shape of the intensity function and badly for other shapes. Even though a simple histogram might already give us a good idea on the shape of the intensity function, it is quite difficult to draw conclusions as to just even the height of intensity peaks for example and this identification is quite important since it makes the difference between getting a very good estimate or a very bad one – a situation that seems quite undesirable in not just our present context. It also seems a bit inadequate to have a condition on the intensity function while estimating the period: as already mentioned, having an estimator for the period is crucial to being able to produce good estimators for the intensity function; it seems thus very unnatural to have such a strong condition on the shape of the intensity function given that a good estimate for the intensity function could only have been obtained from a good estimate for the period.

The other, nonparametric, estimators that are already known for the period – whose list we hope is complete – seem to follow a rather simplistic take on the problem, by relying mostly on the fact that the number of events per period should be approximately constant. The criterion is then that the number of events over the elements of a partition in time intervals corresponding to the correct value for the period should minimise the squared difference with the average number of events over all time intervals. While of course sane, this approach seems to fail at taking advantage of the fact that in the background we have an infinitely dimensional parameter – the intensity function – that is the sole responsible for the pattern that events follow in a trajectory.

While the properties of the estimator presented by Mangku in [17] are theoretically established, the one presented by Bebbington and Zitikis in [15] lack the same sort of support and the authors try only to lay down the track for a theoretical establishment of the estimator. Bebbington and Zitikis do however state that Mangku's estimator is of low practical use since it often seems to produce estimates that correspond to fractions of the period, as they present an example of a simple intensity function that causes this. Also the periodogram estimator of Vere-Jones [11] is mentioned to be of limited use since it is designed for a very specific type of intensity function.

Estimators for the period seem therefore to be in demand especially since they are a fundamental ingredient in the estimation of the intensity function itself – so strong is this need that the common assumption in this context is even that the period is known exactly, an all but realistic expectation however. The importance of knowing the period (or an estimate thereof) is mostly connected with the ability to partition a trajectory into sub intervals of length τ and effectively construct a sample of observations of one cycle of the process, opening the way to the estimation of the intensity function; and one would agree that having a sample or knowing how to construct one is fundamental need in statistics.

It seems thus that a good estimator should have a certain balance. It should not be mostly insensitive, like the present nonparametric estimator, to what happens *within* the period and try to explore the presence of patterns in the trajectories of the process. On the other hand it should also not be too sensitive; we can, for a fixed period, have, in essence, any function within one period of the intensity function and we still expect the estimator to converge to the same value in all of these cases.

This paves the way to the construction of our estimator. The basic idea was to indirectly estimate the period by introducing a value T and estimating instead the difference between T and the period. This extra degree of freedom for the estimator would turn out to be of great importance in helping us to know if we are indeed estimating the period and not one of its multiples.

In a first moment we assume to have several trajectories coming from processes that share the same intensity function. This might be an unrealistic assumption since it would be roughly equivalent to a situation where we would be able to know where the period starts (but not where it ends). These assumptions simplify the situation theoretically but are presented here mainly for the reason that it is simple to relate what is being done by counting events and comparing counts of events, with its deterministic counterparts of integrating and comparing areas. It serves then mainly as an illustration for our approach, an illustration that would be difficult to present directly for our proposed estimator (the second version of the estimator).

By giving up, while constructing the second version of the estimator, on the assumption of a common distribution for the several trajectories we are working with, we do gain something in return though: the intervals whose counts are being compared correspond to different regions of the period of the intensity function; we thus try to take advantage of the fact that we are working with a cyclical function in a more marked way: it is not only that the number of events over the whole period should be close, there are infinitely many intervals where this should happen. By building an estimator that takes this into consideration we are creating a much more demanding criterion function and so hope to obtain better estimates – Mangku also noted this phenomenon in [17, p.105] when he mentions how the variance of its estimator decreases when the intensity function becomes less flat: if the criterion function is less flat, it becomes more difficult for values of the parameter other than the correct one to minimise the criterion function, resulting with a steeper criterion function which, at least in the context of Z-estimators, is connected with lower asymptotic variance.

It was also quite interesting to see that we still have convergence to the intensity function even when we for each fixed θ we are counting events that where produced by different intensities; it suggests that we can always talk about an *average number of events over intervals of a certain fixed length.* It can also be due to the fact that for rational τ and rational T, even though we are working with differently distributed (sub-) trajectories (after partitioning our initial trajectory) we can have a number of possible distributions that is at most the least common multiple between τ and T – which is finite if they are both rational. This should be enough, in our NHPP context, to assure convergence at least pointwise for our event counts.

The stability in the results are quite possibly connected with the fact that the quantities we are comparing in the criterion function are always estimates for integrals that are varying continuously as the argument of the criterion function changes; this is quite good since it dampens the effect that having no constraints on the shape of the intensity function could have. Regardless, continuity is without doubt quite essential in this context; if it were not to hold in a neighbourhood of θ_0 it could even render the criterion function useless.

On a more general note, it might seem strange how it is even possible to identify the period; we have a known mechanism that produces a list of events times; we are under the assumption that there is a certain pattern repeating itself but we have *a priori* no idea of what this pattern is; still, we are able to discover how often the pattern repeats itself. The situation looks particularly grim with respect to the identifiability of the period if we note that literally every periodical function can be distorted to have a certain fixed period; this means that the pattern that we are looking is not only a random realisation of itself but also this pattern can be *any*. The situation looks even worst when we realise that while working with a single trajectory of the process a solution for the problem will most likely involve mixing samples with different distributions.

We are however only interested in quite a small piece of the information that is unknown: the period. The identifiability of the period should then possible since we are using a criterion function that, as already partially mentioned, encompasses for each θ a scaled average difference between the number of events over two specific intervals of a certain fixed length, or if we prefer an average relative difference between the number of events over said intervals; the fact that this should characterise, to great extent, the underlying intensity function is the key to a proper identification of the period; we thus have a parallel: the criterion function contains the complexity of the intensity function while the local minima (or zeros, depending on the version) of the criterion function contain the information regarding the length of the period; while reverse constructing the intensity function from the criterion function would probably be a devious task, recuperating the period from the criterion function is immediate; this seems to be the biggest merit of using this method: it shifts the complexity from identifying the period (what we are interested in) to the identification of the intensity function (a problem we are not concerned with).

Another issue that should be mentioned is the amount of minima that the intensity function has. Given a fixed T, and considering even just $\Theta = [0, T/2]$ we will have a minimum at $T - \tau$ (if $T > \tau$) plus other minima corresponding to other multiples of the period that might be in the interval and also other points at which we might, by chance, also accumulate the same area; even though the theorems M-estimators require there to be only one local minimum in the parameter set we work with, and our estimator, being closely related to one should also need this assumption, the graphs presented throughout the text sometimes show more than one local minimum; we only represent here the full (actually half) of the criterion function for illustrative purposes – it is quite easy to find in practise parameter sets Θ that contain a unique local minima even for very low sample sizes. All of this is also valid for the zeros of the first version of the estimator.

This is, in our opinion where the previous nonparametric estimators, at least in the fashion in which they are presented by their authors, have their greatest fault: the justification that given a proper parameter set Θ the estimator will indeed converge to the desired value τ. The discussion on how to distinguish τ from its multiples seems weak and perhaps even worst, the authors do not discuss how τ can be distinguished from points *other* than its multiples that also minimise the criterion function that they use; one would expect this as the next step in their analysis when they discover example of intensity functions that have these points. Without a careful look at this there is the possibility that we obtain faulty estimates.

For the second version of the estimator these is a simple way, when in the presence of several local minima, and for $\tau < T < 2\tau$ to immediately identify which minimum corresponds to the difference to the multiple of the period: since no other point should verify the equality between the two areas being compared more often than $T - \tau$ (at least in probability), then this local minimum must be the one for which the criterion function takes its lowest value.

We also note that while in [15, 17] the existence of flat portions in their own criterion function poses a serious problem since it does not allow for the identification of a unique estimate; in our case, the existence of the extra parameter T gives us the freedom to overcome situations like this one by simply choosing another value for T.

Even though this was not mentioned explicitly, the establishment of the asymptotic distribution of the estimator would allow us to build confidence intervals for the period and equivalently do hypothesis testing.

The mapping consisting of the results produced by the estimator versus the value of T were shown to be quite effective in helping us identify a good parameter set to work with. Relying on the fact that the period should always be of a larger order than the variability that the estimates have, the steps in this graph will always be evident; anomalies in the graph introduced due to punctual symmetries or relations between the shape of the different parts of the intensity function could also be quite easily identified an posed no great problem as long as we concentrate ourselves on the largest candidate estimate for the period on $[0, T/2]$ since multiples of the period will *always* verify our criteria of the areas being compared matching; steps corresponding to other points will always drop to previous levels.

We also saw that the first version of the estimator was capable of producing more accurate estimates than the second version; so, what we gain from considering a more demanding criterion function does not seem to fully compensate for the loss of a common distribution over the intervals of length T we are using to build the criterion function. It is also quite nice to see that pleasant properties for these functions such as boundedness and continuity hold for both versions of the estimator.

The repeated estimation of the period was also quite useful in giving us more stable results and allowing to test the normality of the results produced by the estimator; for these tests normality was never rejected as we would expect from the theoretical derivations. While in practise this repeated estimation requires a lot of data, or, for a fixed total amount of data implies each estimate will be built upon a small amount of data, techniques such as bootstrap would allow us to make good use of the available data, making this practise not only a possible option but also a recommended one.

Our prediction for the asymptotic variance of the first version of the estimator could also seems to hold. Indeed the variance of the sample obtained by repeated estimation seemed to adjust quite well to the curve obtained directly from the intensity function. There did not seem to be radical differences between the variance for each choice of T; according to our tentative expression for the variance though, this quantity increases quite rapidly as we consider values of T closer to multiples of the period. In our case the amount of data was not an issue since we could simulate any amount of it making the final choice for T boil down to choosing the value that resulted in a lower value for the asymptotic variance of the estimator; however in a less ideal circumstance the underlying trade-off needs to be considered: a smaller value for T will result in a larger sample but might not compensate for the fact that choosing a larger value for T might correspond to a smaller asymptotic variance; a larger choice for T in hopes of taking advantage of a smaller asymptotic variance might not compensate for the fact that will be reducing the size of the sample. The general recommendation would therefore be to use re-sampling techniques to get an idea of the relation between T and the variance of the estimator and then choosing the value of T that results in a lower variance (note that this way we would immediately be making the best use of the available data). Another alternative brings us back to the relation of our problem with the problem of estimating λ; if an estimate for the intensity function is known, we can use it to get an idea of the variance of the estimator as a function of T.

Another variance that deserves our attention is the pointwise variance of the criterion functions. We saw that, as expected, for the first version of the estimator, that this variance increases with θ since we make no attempt of normalising the criterion function in terms of its variance; for the second version of the estimator, however, we see that the sum in the denominator of the criterion function does have some effect in terms of variance: the variance of the criterion function for θ close to θ_0 is close to one and, perhaps most importantly, remains at roughly this level not changing drastically. This would indicate that indeed the difference between the our normalisation factor and the a factor that would normalise the variance of the full criterion function is not big when we are close to θ_0 and thus serves its purpose.

The effect of the changing the average number of events per cycle resulted for the first version of the estimator in the expected way with the variance of the estimates being increases by a factor of $\frac{1}{r}$ for r the ratio of the maximum values for the intensity; other than this there do not seem to by any major differences resulting from the scaling. As for the second version of the estimator the effect seems to extend beyond the variance of the estimates: the variance does seem to behave in the same way as with the first version of the estimator (although reaching higher values); however there seems to be a tendency to overestimate the period that decreases as the average number of events per cycle increases. The reason for this might be related with the fact that we no longer have a sample following a unique distribution. By looking at the effect that the normalisation factor has on the criterion functions if anything it reduces this effect; also the presence of the absolute value does not seem like a likely candidate, leaving the fact that the elements in the sample do not have a common distribution. It was also possible to see how useful the normalisation factor was in strengthening the variation of the criterion function around θ_0; without it would be difficult to identify the local minima.

We could also compare the results that our estimator produces with those of the other existing estimators for the period. The results were quite positive not only compared to the ones obtained by the other estimators but also considering that in all cases the amount of data being used was quite small. Even though the number of events per cycle is of course important in that it adds "resolution" to the criterion function, our estimate depends in essence of the behaviour of said function around θ_0 and so the main factor influencing the results should, and indeed can be seen to be, the number of full cycles contained in the trajectory we are working with; it is therefore quite reassuring that our estimator was able to stay within an acceptable range of the period in 70% of the cases versus just 10% for the best that the other estimators could offer while working with just 5 repetitions of the cycle which in our case meant a sample of size 4.

The results for the periodogram might seem quite poor (and in this case they are indeed) but this is simply a consequence of the shape of the intensity function we were using simply not fitting into the one assumed for this estimator; in fact, the results produced by it are better than the ones for the other estimators for intensity function that simply alternate between a low intensity zone and a high intensity zone within one period – read, intensities that do not have multiple peaks corresponding to different values. For the intensity that was used in this text though it was even difficult to relate the average estimate produced by this estimator and the correct value for period.

The important step of the identifiability of the period could be seen not to be possible only for our particular choice for the intensity function: also with other intensity functions, and without in any way changing the estimator, we were able to find appropriate values for T that allow a proper estimation of the period; this was still true for two "extreme" cases: an intensity function comprised of steps – where the assumption of non *homogeneousness* and therefore the existence of the period holds quite weakly – and when the intensity function almost repeats itself within one period – and the assumption of the period being τ almost fails. It was possible to see here that this identification was easier for the second version of the estimator than for the first version where it was easier for event counts to coincide at points other than θ_0.

The greatest limitation of this estimator (or any estimator for the period for that matter) is that its sample is a set of trajectories whose length has to be of at least one period – it would be difficult to even imagine an estimator that could have as a "sample unit" a trajectory of much smaller length; so, especially for phenomenon with a large period we might be very limited in terms of the amount of data we can collect within a reasonable amount of time be it due to the costs of collecting data or simply due to temporal constraints – common concerns in applied statistics. It is therefore important to make the best use of the data that is available; luckily it is quite possible that there might be a simple way to do this in certain cases. We remind that we are in essence overlapping (while constructing the criterion function) different regions of the intensity function. Imagine now that we have a certain situation that repeats itself (we assume independently) – say for example the volume of persons using different post offices; there is really no reason to believe that the intensity functions of the several post offices are significantly different; there is even less reason to believe that the period should vary from post office to post office. It is then the author's belief that a parallel collection of data from these different sources could collectively produce a valid sample for estimating the common period – the criterions function should be, for fixed T simply the mean of the several criterion function corresponding to the several sources. We note that the only issue here is really the independence between the sources (for a fixed, finite amount thereof). Even if the intensity functions are scaled versions of each other there should be no problem since the estimator should work for all of the sources separately. Under the weaker assumption that the intensity functions really have different shapes mixing the samples should still work as long as we assume that there is a uniform, finite, bound on all intensities – we note that we have already established that the estimator works if each trajectory of length T has a different distributions, and that no restrictions were imposed on the shape of the intensity function. Mixing the samples could possibly introduce other local minima (if the difference between the intensity function of different sources extends beyond just their scale) but if indeed there is a common period we will preserve the location of the local minimum and be in no worst situation when it comes to find the local minimum that corresponds to the period.

Another way to do this would simply be (again under the assumption of independence between the sources) to simple overlap the (unsplit) trajectories from the several sources. We would be adding the NHPP and thus get a single NHPP with an intensity function that would be the sum of the individual intensity function: if there is a common period it will be preserved. In this case it is simpler to the effect that this would have on the variance of our estimate; if both sources would have the same intensity function (including scale) then the variance would drop to half.

While this procedure does not make for a more economical solution for the estimation, it surely allows us to overcome – or at least minimise – the effect of the more constraining physical limitations of data collection. The establishment of the theoretical properties in this case should be no more than a direct extension of our present case under the assumption of independence between the sources and the existence of a common period.

We close by noting that in our analysis we did not come across a phenomenon mentioned in [16]: when doing M-estimation based on discrete data, it is possible for the density of the asymptotic distribution of the estimator not to be normal but have density corresponding to the concatenation of two halves of the densities of normal distributions with the same mean but different variances. It is quite possible that this occurred since our criterion functions, albeit being step functions, converged to a continuous function.

BIBLIOGRAPHY

[1] Lewis, P.A.W. (1972) *Recent results in the statistical analysis of univariate point processes*, in P.A.W. Lewis (ed.), Stochatic Point Processes, pp.1-54 Wiley: New York, 1972.

[2] Ridley, M. (1995) *The Red Queen: Sex and the Evolution of Human Nature*, Penguin Books.

[3] van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press.

[4] Rorabaugh, C.B. (1998) *DSP Primer*, McGraw-Hill Professional, 1^{st} edition.

[5] Ross, S. (2003) *Introduction to Probability Models – Eight Edition*, Academic Press.

[6] Ross, S. (2006) *Simulation – Fourth Edition*, Academic Press.

[7] Williams, D (1991) *Probabilities with martingales*, Cambridge University Press.

[8] Irwin, J.O. (1937) *The frequency distribution of the difference between two independent variates following the same Poisson distribution.*, Journal of the Royal Statistical Society: Series A 100 (3): 415416. [Via http://en.wikipedia.org/wiki/Skellam_distribution – retrieved 2 - 11 - 2008]

[9] Wayne, A. (1946) *Inequalities and inversions of order*, Scripta Mathematica 12 (2): 164169. [Via http://en.wikipedia.org/wiki/Rearrangement_inequality – retrieved 16 - 10 - 2008]

[10] Lewis, P.A.W. (1970) *Remarks on the theory, computation and application of the spectral analysis of series of events*, J. Sound Vib. vol. 12 pp. 353-375.

[11] Vere-Jones, D. (1982) *On estimation of frequency in point-process data*, J. Appl Prob 19A:383–394. [Via http://en.wikipedia.org/wiki/Central_limit_theorem – retrieved 2 - 11 - 2008]

[12] Le Cam, L. (1986) *The central limit theorem around 1935*, Statistical Science 1:1, 7891.

[13] Helmers, R. and Mangku, I.W. (2003) *On estimating the period of a cyclic Poisson process.* In: Mathematical statistics and applications: Festschrift for Constance van Eeden. Institute of Mathematical Statistics, Beachwood, pp.345-356.

[14] Helmers, R., Wang, Q. and Zitikis, R. (2007) *Confidence regions for the intensity function of a cyclical Poisson process.* Statistical Inference for Stochastic Processes, Springer, pp –.

[15] Bebbington, M. and Zitikis, R. (2004) *A Robust Heuristic Estimator for the Period of a Poisson Intensity Function* Methodology And Computing In Applied Probability, Volume 6, Number 4, December 2004 , pp. 441-462(22), Springer.

[16] Simpson, D.G., Carroll, R.J., and Ruppert, D. (1987) *M-estimation for discrete data: Asymptotic distribution theory and implications*, Ann. Statist., 15, 657669.

[17] Mangku, I.W (2001) *Estimating the intensity function of a cyclic poisson process*, Ph.D. Thesis, University of Amsterdam, Amsterdam.