The Shark Function – Asymptotic Behavior of the Filtered Derivative for Point Processes in Case of Change Points

MICHAEL MESSE, GABY SCHNEIDER

Institute for Mathematics
Johann Wolfgang Goethe University
Robert-Mayer-Str. 10
60325 Frankfurt (Main), Germany

corresponding author: Gaby Schneider, schneider@math.uni-frankfurt.de

Abstract

This paper contributes to the statistical theory of change point detection in point processes on the line. We study piecewise renewal processes with changes in the rate in adjacent sections. The points in time where the rate changes are termed change points. A common approach for the detection and localization of the change points is the analysis of a filtered derivative process.

We deduce a weak process convergence result for the filtered derivative process in scenarios of change points and thus extend convergence results obtained for the null hypothesis of no change in the rate. Our result states that the shape of the filtered derivative process reminds on a shark’s fin in the neighborhood of a change point. It is used to give bounds for the detection probability of change points. For practical application, we provide estimators for the point process parameters and discuss the distortion of the filtered derivative process that uses these estimators.
1 Introduction

In this paper we contribute to the statistical theory of change point detection, which aims at the detection of structural breaks (so called change points) in time series. A common approach is the analysis of moving sum statistics, compare the textbooks of Brodsky and Darkhovsky (1993); Basseville and Nikiforov (1993); Csörgő and Horváth (1997). Given a time series, its properties of interest are locally analyzed by successively evaluating an auxiliary statistic in a moving window manner. Thereby, the statistic is constructed which shows extreme values when the associated analyzed section overlaps a change point. Here, we study the so called filtered derivative process, where for each point in time the events in two adjacent windows are compared.

Among others, classical change point models such as (Gaussian) sequences of observations (cmp. e.g., Yao and Davis, 1986; Gombay and Horváth, 1990; Horváth, 1993; Husková and Slaby, 2001), linear regression models (cmp. e.g., Horváth, 1993; Horváth et al., 2007; Husková and Kirch, 2012), autoregressive models (cmp. e.g., Davis et al., 1995; Husková et al., 2007, 2008) have extensively been studied. However, results in the setting of point processes on the line are rare (cmp. Kendall and Kendall, 1980; Csörgő and Horváth, 1987; Steinebach and Zhang, 1993; Messer et al., 2014).

In this paper, we focus on renewal processes on the positive line. In an asymptotic scenario, where the total observation horizon grows linearly with the window size, the filtered derivative process was shown to converge to a functional of a standard Brownian motion, given there is no change point in the underlying renewal process (Steinebach and Eastwood, 1995; Messer et al., 2014).

Here, we study the behavior of the filtered derivative process in the scenario of changes in the rate of the underlying point process, which is specified as a piecewise renewal process (section 2). In section 3 we introduce the filtered derivative process for which we show weak process convergence. In the neighborhood of a change point, the process fluctuates around a deterministic function that has the shape of a shark’s fin, and is therefore called the shark function. We use the convergence result in order to give bounds for the detection probability of the change points. Section 4 considers an additional distortion term that results in practical application when the point process parameters need to be estimated. We propose an estimator for the process parameters and discuss the resulting process convergence after plugging the estimator into the filtered derivative process.

2 Piecewise Renewal Processes

We write a point process Φ on the positive line as an increasing sequence of events $0 < S_1 < S_2 < S_3 < \cdots$, where $S_j$ denotes the occurrence time of the $j$-th event, for $j = 1, 2, \ldots$ Alternatively, Φ is determined by its life times $\{\xi_j\}_{j \geq 1}$, where

\[ \xi_1 = S_1 \quad \text{and} \quad \xi_j = S_j - S_{j-1}, \quad \text{for} \quad j = 2, 3, \ldots \]  

or by the counting process $(N_t)_{t \geq 0}$, where

\[ N_t = \max\{j \geq 1 \mid S_j \leq t\}, \quad t \geq 0, \]  

with the convention $\max\emptyset := 0$. The process Φ is called a renewal process with square integrable life times (RP2) if the associated sequence $\{\xi_j\}_{j \geq 1}$ of life times is a sequence
of positive, independent and identically distributed (i.i.d.) and square-integrable random variables with $\sigma^2 := \text{Var}(\xi_1) > 0$. Prominent candidates for the life time distribution are the exponential-, the gamma-, the Weibull- and the log-normal distribution. For an RP2 $\Phi$ it holds $\mu := \mathbb{E}[\xi_1] = \mathbb{E}[\xi_2]$ and $\sigma^2 = \text{Var}[\xi_1] = \text{Var}[\xi_2]$ for all $j = 1, 2, \ldots$, so we write $\Phi = \Phi(\mu, \sigma^2)$, compare Cox (1962); Cox and Isham (1980); Daley and Vere-Jones (2008). The inverse mean $\mu^{-1}$ is termed the rate of $\Phi$. Processes with at most one change (AMOC) in the rate are constructed as piecewise RP2s as follows.

**Construction 2.1.** Let $T > 0$, $c \in (0, T]$ and $n = 1, 2, \ldots$ At time 0 start two independent RP2s $\Phi_1(\mu_1, \sigma_1^2)$ and $\Phi_2(\mu_2, \sigma_2^2)$ with $\mu_1 \neq \mu_2$ and set

$$\Phi^{(n)} := \Phi^{(n)}(c) := \Phi_1[0, nc] \cup \Phi_2(nc, nT),$$

where $\Phi_1[\alpha, \beta]$ denotes the restriction of $\Phi_1$ to the interval $[\alpha, \beta]$. The resulting sequence of interest is given as $(\Phi^{(n)})_{n \geq 1}$ (cmp. Figure 1).

For every $n = 1, 2, \ldots$ the process $\Phi^{(n)}$ starts as the RP2 $\Phi_1(\mu_1, \sigma_1^2)$ at time zero and jumps into $\Phi_2(\mu_2, \sigma_2^2)$ at the change point $nc$. Note that the construction covers the case of no change point, when $c = T$, and that the variance and the rate can change simultaneously. The value $n$ is required for asymptotic statements throughout this work, which are deduced by letting $n \to \infty$. As one can see from the construction, the total time $nT$ and the location of the change point $nc$ grow linearly in $n$.

Figure 1: Schematical representation of the involved quantities in the one change point case. The change appears at time $nc$. Before $nc$, the resulting process $\Phi^{(n)}$ equals an RP2 $\Phi_1$ with parameters $\mu_1$ and $\sigma_1^2$ and after $nc$, it derives from a second RP2 $\Phi_2$ with parameters $\mu_2$ and $\sigma_2^2$. At time $nt$ the associated left window $(n(t-h), nt]$ entirely refers to $\Phi_1$ but the right window $(nt, n(t+h)]$ overlaps the change point $nc$ and refers to life times from $\Phi_1$ and $\Phi_2$.

Given $T$, all sequences $(\Phi^{(n)})_{n \geq 1}$ from construction 2.1 constitute our model set $\mathcal{M} := \mathcal{M}(T)$. We study elements of that model set and investigate the null hypothesis $H_0 : (\Phi^{(n)})_{n \geq 1} \in \mathcal{M}$ and $c = T$, i.e., there is no change point, vs. $H_A : (\Phi^{(n)})_{n \geq 1} \in \mathcal{M}$ and $c < T$, i.e., there exists a change point.

### 3 The Filtered Derivative Process in Case of a Change Point

In order to test the null hypothesis and to estimate the location of the change point, one common approach is a filtered derivative procedure (Brodky and Darkhovskiy, 1993; Basseville and Nikiforov, 1993; Csörgő and Horváth, 1997). We first define a specific filtered derivative process for the detection of rate changes in renewal processes used also in Steinebach and Eastwood (1995);
Theorem 3.2. We recall a convergence result under $H_0$ and prove an extended convergence result under $H_A$.

Throughout the article, we use the following notation: Let $(\Phi^{(n)})_{n \geq 1} \in \mathcal{M}$. For $T > 0$ let $h \in (0, T/2]$ denote a window size and $\tau_h := [h, T - h]$ an analysis region. Let $(N_t^{(n)})_{t \geq 0}$ denote the counting process corresponding to $\Phi^{(n)}$.

**Definition 3.1.** For $t \in \tau_h$ the filtered derivative process $D^{(n)} := (D_t^{(n)})_{t \in \tau_h}$ is defined as

$$D_t^{(n)} := D_{h,t}^{(n)} := \left( \frac{N_{n(t+h)}^{(n)} - N_{nt}^{(n)}}{s_t^{(n)}} - \frac{N_{nt}^{(n)} - N_{n(t-h)}^{(n)}}{s_t^{(n)}} \right)$$ for $t \in \tau_h$. \(\text{(4)}\)

Thus, $D_t^{(n)}$ compares the number of events in a left window, $N_{n(t+h)}^{(n)} - N_{nt}^{(n)}$, to the number of events in a right window, $N_{nt}^{(n)} - N_{n(t-h)}^{(n)}$ (cmp. Figure 4). The standard deviation $s_t^{(n)} := s_{h,t}^{(n)}$ is given below in equation (3) and is required for process convergence. The process $D^{(n)}$ can indicate changes in the rate because its expectation asymptotically vanishes if the analysis window does not overlap the change point, while systematic deviations from zero are expected when a rate change occurs within the analysis window.

In case of no change point (i.e., $c = T$) weak process convergence of $D^{(n)}$ to a functional of the standard Brownian motion was shown in Steinebach and Eastwood (1993) and Messer et al. (2014) for renewal processes and certain generalizations with respect to variability in the variance. The main results are summarized in the following Lemma in which for the null hypothesis we restrict $\Phi^{(n)}$ to originate from one single RP2 $\Phi_1(\mu_1, \sigma_1^2)$. Let $D[h, T - h]$ denote the set of all càdlàg functions on $[h, T - h]$ and $d_{SK}$ the Skorokhod metric on $D[h, T - h]$.

**Theorem 3.2.** Let $\Phi_1(\mu_1, \sigma_1^2)$ be an RP2, let $c = T$ such that $\Phi^{(n)} = \Phi_1|_{[0,T]}$. Let $(W_t)_{t \geq 0}$ be a standard Brownian motion. Then it holds in $(D[h, T - h], d_{SK})$ as $n \to \infty$

$$\left( \frac{N_{n(t+h)}^{(n)} - N_{nt}^{(n)}}{\sqrt{2nh\sigma_1^2/\mu_1^3}} - \frac{N_{nt}^{(n)} - N_{n(t-h)}^{(n)}}{\sqrt{2nh\sigma_1^2/\mu_1^3}} \right) \configured{t \in \tau_h} \xrightarrow{d} \left( \frac{W_{t+h} - W_t - W_t - W_{t-h}}{\sqrt{2h}} \right) \configured{t \in \tau_h}. \text{(5)}$$

The left hand side equals $D_t^{(n)}$ for $s_t^{(n)} = (2nh\sigma_1^2/\mu_1^3)^{1/2}$, which holds for $c = T$, as we will see below (equation (3)). In case of a change point, $D^{(n)}$ systematically deviates from zero for $t \in (c - h, c + h)$. Therefore, we need an additional centering term in order to prove convergence.

**Definition 3.3.** Let the rescaled filtered derivative process $\Gamma^{(n)} := (\Gamma_t^{(n)})_{t \in \tau_h}$ be defined as

$$\Gamma_t^{(n)} := \Gamma_{h,t}^{(n)} := \left( \frac{N_{n(t+h)}^{(n)} - N_{nt}^{(n)}}{s_t^{(n)}} - \frac{N_{nt}^{(n)} - N_{n(t-h)}^{(n)}}{s_t^{(n)}} \right) - m_t^{(n)}$$ for $t \in \tau_h$. \(\text{(6)}\)

while for $t \in \tau_h$ the expectation function $m_t^{(n)} := (m_t^{(n)})_{t \in \tau_h}$ is zero for $|t - c| > h$ and equals

$$m_t^{(n)} := m_{h,t}^{(n)}(c) := n (1/\mu_2 - 1/\mu_1) (h - |t - c|) \text{ for } |t - c| \leq h \text{ (see Figure 4 A, C).} \text{(7)}$$
Analogously to Lemma \[3.2\], the variance \((s_t^{(n)})^2 := \left((s_t^{(n)})^2\right)_{t \in \tau_h}\) is given by \(2nh\sigma_t^2/\mu_t^3\) for \(t < c - h\), by \(2nh\sigma_t^2/\mu_t^3\) for \(t > c + h\), and by a linear interpolation (see Figure 2 B, D)

\[
(s_t^{(n)})^2 := \left((s_t^{(n)})^2\right) := n \left((t + h - c)\sigma_t^2/\mu_t^3 + (c - (t - h))\sigma_t^2/\mu_t^3\right), \quad \text{for } |t - c| \leq h. \tag{8}
\]

Figure 2: Representation of the expectation function \(m_t\) (A, C) and the variance function \(s_t^2\) (B, D) in case of a change point at \(c\), according to Definition 3.3. The expectation \(m_t\) vanishes outside \([c - h, c + h]\) and takes its extreme, \(m_c = (1/\mu_2 - 1/\mu_1)nh\), at \(c\). The function \(m_t\) is non-negative if the rate increases (A) and non-positive if the rate decreases (C). The variance function \(s_t^2\) equals \(2nh(\sigma_t^2/\mu_t^3)\) for \(t < c - h\) and \(2nh(\sigma_t^2/\mu_t^3)\) for \(t > c + h\) and is linearly interpolated in \([c - h, c + h]\) (B, D). Superscripts \(n\) are omitted here for convenience in the notation of \(m\) and \(s\).

We now show weak convergence of \(\Gamma^{(n)}\) in Skorokhod topology to a limit process \(L\) in the general setting of a change point.

**Proposition 3.4.** Let \(\Phi_1(\mu_1, \sigma_1^2)\) and \(\Phi_2(\mu_2, \sigma_2^2)\) be independent RP2s with \(\mu_1 \neq \mu_2\). Let \(c \in (0, T]\) be a change point, so that the sequence \((\Phi^{(n)})_{n \geq 1}\) results from \(\Phi_1\) and \(\Phi_2\) according to Construction 2.7, and let \(\Gamma^{(n)}\) be the associated rescaled filtered derivative process. Let \((W_t)_{t \geq 0}\) be a standard Brownian motion, and the limit process \(L := (L_t)_{t \in \tau_h}\) be given as

\[
L_t := L_{h,t}(c) := \begin{cases} 
\frac{(W_{t+h}-W_t)-(W_t-W_{t-h})}{\sqrt{2h}}, & \text{if } |t - c| > h, \\
\sqrt{\sigma_t^2/\mu_t^3}(W_{t+h}-W_t)+\sqrt{\sigma_t^2/\mu_t^3}(W_t-W_{t-h}) & \text{if } c - h \leq t \leq c, \\
\sqrt{\sigma_t^2/\mu_t^3}(W_{t+h}-W_t)-(W_t-W_{t-h}) & \text{if } c < t \leq c + h.
\end{cases} \tag{9}
\]

Then it holds in \((D[h,T-h],d_{SK})\) as \(n \to \infty\)

\[
\Gamma^{(n)} \overset{d}{\to} L.
\]

Note that Lemma 3.2 is a special case by setting \(c = T\), because for all \(t \in \tau_h\), we obtain \(m_t^{(n)} = 0\), \(s_t^{(n)} = (2nh\sigma_t^2/\mu_t^3)^{1/2}\), \(L_t = [(W_{t+h} - W_t) - (W_t - W_{t-h})]/(2h)^{1/2}\) and \(\Gamma_t^{(n)}\) equal to the left hand side in equation (6).
Proof of Proposition 3.4: Let \( \{\xi_{1,j}\}_{j \geq 1}, \{\xi_{2,j}\}_{j \geq 1} \) and \( \{\xi^{(n)}_{j}\}_{j \geq 1} \) denote the sequences of life times that correspond to \( \Phi_1, \Phi_2 \) and to the compound process \( \Phi^{(n)} \), respectively (equation (1)). Analogously, let \( (N_{1,t})_{t \geq 0}, (N_{2,t})_{t \geq 0} \) and \( (N^{(n)}_{t})_{t \geq 0} \) denote the counting processes that correspond to \( \Phi_1, \Phi_2 \) and to \( \Phi^{(n)} \), respectively (equation (2)). Let \( (W_{1,t})_{t \geq 0} \) and \( (W_{2,t})_{t \geq 0} \) be independent standard Brownian motions.

For \( i = 1, 2 \) let the rescaled random walk \( (X^{(n)}_{i,t})_{t \geq 0} \) and the rescaled counting process \( (Z^{(n)}_{i,t})_{t \geq 0} \) concerning \( \Phi_i \) be given as

\[
X^{(n)}_{i,t} := \frac{1}{\sigma_i \sqrt{n}} \sum_{j=1}^{[nt]} (\xi_{i,j} - \mu_i) \quad \text{and} \quad Z^{(n)}_{i,t} := \frac{N_{i,nt} - nt/\mu_i}{\sqrt{n\sigma_i^2/\mu_i^3}},
\]

for \( t \geq 0 \). According to Donsker’s theorem, we find in \( (D[0,\infty), d_{SK}) \) as \( n \to \infty \) that

\[
(X^{(n)}_{i,t})_{t \geq 0} \xrightarrow{d} (W_{i,t})_{t \geq 0} \quad \text{for} \quad i = 1, 2,
\]

implying weak convergence of \( (Z^{(n)}_{i,t})_{t \geq 0} \), i.e., it holds in \( (D[0,\infty), d_{SK}) \) as \( n \to \infty \) that

\[
(Z^{(n)}_{i,t})_{t \geq 0} \xrightarrow{d} (W_{i,t})_{t \geq 0} \quad \text{for} \quad i = 1, 2,
\]

as stated in [Billingsley (1999), Thm. 14.6].

We use a different scaling and set

\[
\tilde{Z}^{(n)}_{i,t} := \frac{N_{i,nt} - nt/\mu_i}{s^{(n)}_{i,t}}, \quad t \geq 0,
\]

where \( s^{(n)}_{i,t}, t \in [0,\infty) \) is given in Definition 3.3. Then for \( i = 1, 2 \), we find in \( (D[0,\infty), d_{SK}) \) for \( n \to \infty \)

\[
\left( \tilde{Z}^{(n)}_{i,t} \right)_{t \geq 0} \xrightarrow{d} \left( \frac{\sqrt{\sigma_i^2/\mu_i^3}}{s^{(1)}_{i,t}} W_{i,t} \right)_{t \geq 0}
\]

because \( \left( \sqrt{n\sigma_i^2/\mu_i^3}/s^{(n)}_{i,t} \right) = \left( \sqrt{\sigma_i^2/\mu_i^3}/s^{(1)}_{i,t} \right) \) is continuous in \( t \) and does not depend on \( n \).

Let now \( (\tilde{Z}^{(n)}_{1,t})_{t \geq 0} \) and \( (\tilde{Z}^{(n)}_{2,t})_{t \geq 0} \) denote the processes derived from \( \Phi_1 \) and \( \Phi_2 \), respectively. Due to independence of \( \Phi_1 \) and \( \Phi_2 \), we obtain joint convergence in \( (D[0,\infty) \times D[0,\infty), d_{SK} \otimes d_{SK}) \) for \( n \to \infty \)

\[
\left( \left( \tilde{Z}^{(n)}_{1,t} \right)_{t \geq 0}, \left( \tilde{Z}^{(n)}_{2,t} \right)_{t \geq 0} \right) \xrightarrow{d} \left( \left( \frac{\sqrt{\sigma_1^2/\mu_1^3}}{s^{(1)}_{i,t}} W_{1,t} \right)_{t \geq 0}, \left( \frac{\sqrt{\sigma_2^2/\mu_2^3}}{s^{(2)}_{i,t}} W_{2,t} \right)_{t \geq 0} \right).
\]

We observe the continuous map \( \varphi : (D[0,\infty) \times D[0,\infty), d_{SK} \otimes d_{SK}) \to (D[h, T - h], d_{SK}) \) given by

\[
((f(t))_{t \geq 0}, (g(t))_{t \geq 0}) \xrightarrow{\varphi} \left( (f(t + h) - f(t)) - (f(t) - f(t - h))1_{[h,c-h)}(t) + (g(t + h) - g(c)) + (f(c) - f(t)) - (f(t) - f(t - h))1_{[c-h,c)}(t) \right. \\
+ (g(t + h) - g(c)) - (g(t) - g(c)) - (f(c) - f(t - h))1_{[c,c+h)}(t) \left. \\
+ (g(t + h) - g(t)) - (g(t) - g(t - h))1_{[c+h,T-h)}(t) \right)_{t \in \tau_n},
\]
Thus, it remains to be shown that

\[ \varphi \left( \left( Z_{1,t}^{(n)} \right)_{t \geq 0}, \left( Z_{2,t}^{(n)} \right)_{t \geq 0} \right) \xrightarrow{d} \varphi \left( \left( \frac{\sqrt{\sigma_1^2/\mu_1}}{s_t^{(1)}} W_{1,t} \right)_{t \geq 0}, \left( \frac{\sqrt{\sigma_2^2/\mu_2}}{s_t^{(1)}} W_{2,t} \right)_{t \geq 0} \right). \]

Thus, it remains to be shown that

\[ \left( \Gamma_t^{(n)} \right)_{t \in \mathbb{T}_n} = \varphi \left( \left( \tilde{Z}_{1,t}^{(n)} \right)_{t \geq 0}, \left( \tilde{Z}_{2,t}^{(n)} \right)_{t \geq 0} \right), \quad (12) \]

\[ (L_t)_{t \in \mathbb{T}_n} \sim \varphi \left( \left( \frac{\sqrt{\sigma_1^2/\mu_1}}{s_t^{(1)}} W_{1,t} \right)_{t \geq 0}, \left( \frac{\sqrt{\sigma_2^2/\mu_2}}{s_t^{(1)}} W_{2,t} \right)_{t \geq 0} \right), \quad (13) \]

where \( \sim \) denotes equality in distribution. In order to show (12) and (13) we differentiate the four cases \( t \in [h,c-h) \), \( t \in [c-h,c) \), \( t \in [c,c+h) \) and \( t \in [c+h,T-h] \).

**Derivation of (12):**

**Case** \( t < c-h \):

\[ \varphi \left( \left( \tilde{Z}_{1,t}^{(n)} \right)_{t \geq 0}, \left( \tilde{Z}_{2,t}^{(n)} \right)_{t \geq 0} \right) = \frac{(N_{1,n(t+h)} - N_{1,nt}) - (N_{1,nt} - N_{1,n(t-h)})}{s_t^{(n)}} = \Gamma_t^{(n)}. \]

For \( t \geq c+h \) we obtain analogous results by exchanging subscripts. For \( t \in [c-h,c) \) we obtain

\[ \varphi \left( \left( \tilde{Z}_{1,t}^{(n)} \right)_{t \geq 0}, \left( \tilde{Z}_{2,t}^{(n)} \right)_{t \geq 0} \right) \bigg|_{t} = \frac{(N_{2,n(t+h)} - N_{2,nt}) + (N_{1,nt} - N_{1,n(t-h)} - n \left( \frac{(t+h)-c}{\mu_2} - \frac{(t+h)-c}{\mu_1} \right))}{s_t^{(n)}} = \Gamma_t^{(n)} . \]

Analogously, we obtain \( c \leq t < c+h \), which proves (12).

**Derivation of (13):**

For \( t < c-h \) we obtain

\[ \varphi \left( \left( \frac{\sqrt{\sigma_1^2/\mu_1}}{s_t^{(1)}} W_{1,t} \right)_{t \geq 0}, \left( \frac{\sqrt{\sigma_2^2/\mu_2}}{s_t^{(1)}} W_{2,t} \right)_{t \geq 0} \right) \bigg|_{t} = \frac{(W_{1,t+h} - W_{1,ct}) - (W_{1,ct} - W_{1,t-h})}{\sqrt{2h}} = L_t. \]

The same holds for \( t \geq c+h \) with the subscript exchanged. In the case \( c-h \leq t < c \) we
obtain
\[
\varphi\left(\frac{\sqrt{\sigma_2^2/\mu^2}}{s_{(1)}(t)^2} W_{1,t} \right)_{t \geq 0}, \left(\frac{\sqrt{\sigma_2^2/\mu^2}}{s_{(1)}(t)^2} W_{2,t} \right)_{t \geq 0}\right|_t = \frac{\sqrt{\sigma_2^2/\mu^2}(W_{2,t+h} - W_{2,c}) + \sqrt{\sigma_2^2/\mu^2}|(W_{1,c} - W_{1,t}) - (W_{1,t} - W_{1,t-h})|}{s_{(1)}(t)} = L_t. \tag{15}
\]

Analogously, we obtain for \(c \leq t < c + h\)
\[
\varphi\left(\frac{\sqrt{\sigma_2^2/\mu^2}}{s_{(1)}(t)^2} W_{1,t} \right)_{t \geq 0}, \left(\frac{\sqrt{\sigma_2^2/\mu^2}}{s_{(1)}(t)^2} W_{2,t} \right)_{t \geq 0}\right|_t = \frac{\sqrt{\sigma_2^2/\mu^2}|(W_{2,t+h} - W_{2,c}) - (W_{2,t} - W_{2,c})| + \sqrt{\sigma_2^2/\mu^2}(W_{1,c} - W_{1,t-h})}{s_{(1)}(t)} = L_t. \tag{16}
\]

Let now \((W_t)_{t \geq 0}\) be a standard Brownian motion, i.e., \((W_t)_{t \geq 0} \sim (W_{1,t})_{t \geq 0} \sim (W_{2,t})_{t \geq 0}\). The process defined in (14), (15) and (16) has continuous sample paths and is given as a function of increments of disjoint intervals of the processes \((W_{1,t})_{t \geq 0}\) and \((W_{2,t})_{t \geq 0}\). Therefore, we can omit the subscripts one and two in (14), (15) and (16) and obtain a process that has continuous sample paths and the same distribution as the former one. By omitting the subscripts, we obtain the limit process \(L\) as defined in equation (9), which completes the proof of Proposition 3.4. \(\square\)

Note that Proposition 3.4 refers to RP2s, but that the proof applies directly to a larger class of renewal processes with a certain degree of variability in the variance defined in Messer et al. (2014).

### 3.1 The Shark Function

Proposition 3.4 states that asymptotically the following equality in distribution holds
\[
D^{(n)} \sim \Lambda^{(n)} + L, \quad \text{with} \quad \Lambda^{(n)} = m^{(n)}/s^{(n)}. \tag{17}
\]

In order to understand the process \(D^{(n)}\), we investigate \(\Lambda^{(n)}\). The expectation function \(m^{(n)}\) has the shape of a hat (Figure 2 A), but due to the normalization, the function \(\Lambda^{(n)}\) resembles a shark’s fin and is therefore called here the shark function. We show examples of such shark functions in Figure 3 and give a proof in Lemma 3.5.

Equation (18) in Lemma 3.5 states that the shark function takes its largest deviation from zero at time \(c\). If \(m^{(n)} \geq 0\) and \(s^{(n)}\) increasing, the shark is heading west (Figure 3 A, equation (14)), whereas in case of (20), the shark is heading east (Figure 3 B). For \(m^{(n)} \leq 0\) analogous relations hold, the shark is heading in the same directions, but turned upside down (Figure 3 C and D). Note also that if the standard deviation \(s^{(n)}\) is constant over time, the shark function \(\Lambda^{(n)}\) has a hat shape, i.e., is piecewise linear.

**Lemma 3.5.** For \(T > 0\) and \(c \in (0,T)\), \(h \in (0,T/2]\) and \(t \in \tau_h\) let \(m^{(n)}\) and \(s^{(n)}\) be as in Definition 3.3 and \(\Lambda^{(n)} = m^{(n)}/s^{(n)}\). Then \(\Lambda^{(n)}\) is continuous, with
\[
\arg \max \left\{ \frac{\Lambda^{(n)}_t}{t \in \tau_h} \right\} = c. \tag{18}
\]
For $t \notin (c-h, c+h], m^{(n)} = 0$ and thus, $\Lambda^{(n)} = 0$. For $t \in (c-h, c+h]$, we separate four cases: For $m^{(n)} \geq 0$ and $s^{(n)}$ increasing (Figure 3 A),

$$\left(\Lambda_t^{(n)}\right)_{t \in B} \text{ is } \begin{cases} \text{concave and increasing,} & \text{if } B = [c-h, c], \\ \text{convex and decreasing,} & \text{if } B = (c, c+h]. \end{cases} \tag{19}$$

For $m^{(n)} \geq 0$ and $s^{(n)}$ decreasing (Figure 3 B),

$$\left(\Lambda_t^{(n)}\right)_{t \in B} \text{ is } \begin{cases} \text{convex and increasing,} & \text{if } B = [c-h, c], \\ \text{concave and decreasing,} & \text{if } B = (c, c+h]. \end{cases} \tag{20}$$

For $m^{(n)} \leq 0$, expressions (19) and (20) hold true, but with ‘convex’ and ‘concave’ as well as ‘increasing’ and ‘decreasing’ exchanged.

Further, because $m_t^{(n)}$ is of order $nh$ and $s_t^{(n)}$ is of order $(nh)^{3/2}$ for $|t-c| < h$, we find that $\Lambda_t^{(n)}$ is of order $(nh)^{1/2}$ for $|t-c| < h$. 

Figure 3: Analysis of the shark function $\Lambda_t$ (solid), for the case of a change point at $c$. The dotted line marks the scaled hat function $m_t/s_c$. The shape of the shark function depends on the structure of the expectation function $m_t$ and the standard deviation function $s_t$. For $m_t \geq 0$ and $s_t^2$ increasing, the shark is going west (A). For $m_t \geq 0$ and $s_t^2$ decreasing, the shark is heading east (B). For $m_t \leq 0$, the shark is swimming upside down and is oriented towards the same directions (C and D). The expectation and standard deviation functions refer to point processes whose life times before the change point are i.i.d. $\Gamma(p_1, \lambda_1)$ distributed and those after the change point are i.i.d. $\Gamma(p_2, \lambda_2)$ distributed. The parameters are given in the upper table. Further parameters are $T = 1000$, $c = 500$, $h = 150$ and $n = 1$. Superscripts $(n)$ are omitted for convenience.
Proof of Lemma 3.5: Continuity is clear because both the numerator and the denominator are continuous. Further, we deduce the case \( m_t(n) \geq 0 \) and \( s_t(n) \) increasing. For \( t \in [c-h, c] \), both functions \( m_t(n) \) and \( s_t(n) \) are monotone increasing in \( t \). While \( m_t \) if of order \( t^{1/2} \), see equations (7) and (8). Thus, the shear function \( \Lambda_t(n) \) is monotone increasing and of order \( t^{1/2} \), and therefore describes a concave function for \( t \in [c-h, c] \). For \( t \in (c, c+h] \), \( m_t(n) \) is monotone decreasing and of order \( t \), so that \( \Lambda_t(n) \) is monotone decreasing of order \( t^{1/2} \), which describes a convex function. The other cases follow by similar arguments. In particular, it follows that \( |\Lambda_t(n)| \) is maximized for \( t = c \). □

The fact that \( \Lambda(n) \) takes its maximal deviation from zero at the change point \( c \) can be used for change point detection and for a rough evaluation of the detection probability of a change point, as we show in section 3.1.1.

3.1.1 Detection Probability in Change Point Estimation

In practice, the null hypothesis of no rate change is rejected if the filtered derivative process \( D^{(1)} \) exceeds a rejection threshold \( Q \), i.e., if \( \max_{t \in \tau_n} |D^{(1)}_t| > Q \). The threshold \( Q \) can be derived by Monte Carlo simulation as a quantile of \( \max_{t \in \tau_n} |L_t| \) or by Bootstrap procedures, compare e.g. [Messer et al. 2014]. If the null hypothesis is rejected, one makes use of the observation that \( D^{(1)} \) is expected to take its maximum at time \( c \), such that in the one change point case, the estimate of \( c \) is given as \( \hat{c} := \arg \max_{t \in \tau_n} |D^{(1)}_t| \). In case of multiple change points \((0 = c_0 < c_1, \ldots, < c_k < c_{k+1} = T)\) successive argmax-type estimation methods are applied. Lemma 3.5 can be generalized if the change points are sufficiently far apart, i.e. if \( c_{i+1} - c_i \geq 2h \). In this case, the corresponding function \( \Lambda(n) \) describes multiple, successive shark functions and an analogous convergence result holds. If change points lie closer together, more technical effort is necessary in order to formulate an appropriate convergence result. For further results on argmax estimation see e.g. [Carlstein 1988; Dümbgen 1991; Antoch and Hůsková 1994; Antoch et al. 1997; Bertrand 2000; Bertrand et al. 2011; Messer et al. 2014; Muhsal 2013].

According to the construction \( D^{(n)} = \Lambda(n) + \Gamma(n) \), we find asymptotically

\[
D^{(n)}_c \sim \Lambda_c(n) + L_c \sim N \left( \frac{1/\mu_2 - 1/\mu_1}{(\sigma_2/\mu_2^3 + \sigma_1/\mu_1^3)^{1/2}} (nh)^{1/2}, 1 \right)
\]

for one change point \( c \in \tau_n \), according to Proposition 3.3 and equations (7) and (8). Thus, for a given rejection threshold \( Q \), we can asymptotically bound the detection probability of the change point as

\[
P \left( \max_{t \in \tau_n} |D^{(n)}_t| > Q \right) \geq 1 - F \left( Q - \frac{|1/\mu_2 - 1/\mu_1|}{(\sigma_2/\mu_2^3 + \sigma_1/\mu_1^3)^{1/2}} (nh)^{1/2} \right),
\]

(21)

where \( F \) denotes the cumulative distribution function of the standard normal distribution. This can be seen by separate consideration of rate increases and rate decreases. For \( \mu_2^{-1} > \mu_1^{-1} \), we find \( m_c > 0 \) and \( D_c > 0 \), such that \( P(\max_{t \in \tau_n} |D^{(n)}_t| > Q) \geq P(D^{(n)}_c > Q) \). Analogously for \( \mu_2^{-1} < \mu_1^{-1} \), \( P(\max_{t \in \tau_n} |D^{(n)}_t| > Q) \geq P(D^{(n)}_c < -Q) \).
4 The Distortion – Estimation of Process Parameters

The definition of the filtered derivative process \( D^{(n)} \) as in equation (11) relies on the assumption that the theoretical standard deviation \( s^{(n)} \) is known. However, this function depends on the point process parameters \( \mu_1, \mu_2, \sigma_1^2 \) and \( \sigma_2^2 \), which typically need to be estimated in practical application. Here, we therefore use an estimator \( \hat{s}^{(n)} = (\hat{s}_{h,t})_{t \in \tau_h} \) proposed in Messer et al. (2014) that consistently estimates the function \( s^{(n)} \) under the null hypothesis. The estimator \( \hat{s}^{(n)} \) is given by

\[
\left( \hat{s}^{(n)}_t \right)^2 := \left( \hat{s}_{h,t} \right)^2 := \left( \frac{\hat{\sigma}_2^2(nh, nt) + \hat{\sigma}_3^2(nh, nt)}{\hat{\mu}_2^2(nh, nt)} \right) nh \quad \forall t \in \tau_h,
\]

where \( \hat{\mu}_2(nh, nt) \) and \( \hat{\sigma}_2^2(nh, nt) \) denote the empirical mean and variance of all life times whose corresponding point events lie in the right window \( (nt, nt + h) \) or the left window \( (nt - h, nt) \), respectively. If no life times can be found in the respective intervals, the estimators are set to zero.

Replacing \( s_t^{(n)} \) with this estimate \( \hat{s}_t^{(n)} \), we study the convergence of a new process defined as

\[
G^{(n)}_t := \frac{N(n)_{nt+h} - N(n)_{nt}}{\hat{s}_t^{(n)}} - \frac{N(n)_{nt} - N(n)_{nt+h}}{\hat{s}_t^{(n)}} \quad \text{for} \quad t \in \tau_h.
\]

Under the null hypothesis of no change point (i.e., \( c = T \)), the following convergence result is provided in Messer et al. (2014).

**Lemma 4.1.** Let \( \Phi_1(\mu_1, \sigma_1^2) \) be an RP2, let \( c = T \), such that \( \Phi^{(n)} = \Phi_1|_{(0, nT]} \). Then, we have in \( (D[h, T - h], d_{SK}) \) as \( n \to \infty \)

\[
G^{(n)}_t \xrightarrow{d} \left( \frac{(W_{t+h} - W_t) - (W_t - W_{t-h})}{\sqrt{2h}} \right)_{t \in \tau_h}.
\]

The proof of Lemma 4.1 uses the almost sure convergence

\[
\left( \frac{s_t^{(n)}}{\hat{s}_t} \right)_{t \in \tau_h} = \left( \frac{(2nh\sigma_2^2/\mu_1^3)^{1/2}}{\hat{s}_t} \right)_{t \in \tau_h} \to (1)_{t \in \tau_h},
\]

which holds in \( (D[h, T - h], d_{||}) \) as \( n \to \infty \), where \( d_{||} \) denotes the metric induced by the supremum norm. In particular, this convergence states the strong consistency of the estimator \( \hat{s}^{(n)} \) under the null hypothesis. Using this consistency and Lemma 3.2, Lemma 4.1 can be shown by applying Slutsky’s theorem. Compare Messer et al. (2014) for details of (25).

In the general case of a point process, an analogous convergence result for the estimator \( \hat{s}^{(n)} \) holds true. One can show that in \( (D[h, T - h], d_{||}) \) it holds almost surely as \( n \to \infty \) that

\[
\left( \frac{\hat{s}_t^{(n)}}{s_t} \right)_{t \in \tau_h} \to \left( \frac{s_t^{(1)}}{s_t} \right)_{t \in \tau_h} =: (\Delta_t)_{t \in \tau_h},
\]

where \( \hat{s}_t \) describes the limit behavior of \( \hat{s}^{(n)} \) under the alternative. If \( c \in (0, T) \), \( \hat{s}_t \) does not equal \( s_t^{(n)} \), but leads to a distortion. The limit function \( \Delta := (\Delta_t)_{t \in \tau_h} \) is therefore termed here...
the distortion function. It is continuous and depends on the process parameters $\mu_1, \mu_2, \sigma_1^2$ and $\sigma_2^2$ (see Figure 1B for examples). More precisely, $\hat{s}_t$ is given as

$$\hat{s}_t^2 := (s_{h,t})^2 := \left( \frac{\sigma_{r_1}^2(h,t)}{\mu_{r_1}^3(h,t)} + \frac{\sigma_{r_2}^2(h,t)}{\mu_{r_2}^3(h,t)} \right) h \quad \forall t \in \tau_h,$$

with $\mu_{r_1}(h,t) = \mu_1$ for $t \leq c - h$, $\mu_{r_1}(h,t) = \mu_2$ for $t > c$, and

$$\mu_{r_1}(h,t) = h\mu_1\mu_2/((c - t)\mu_2 + (t + h - c)\mu_1) \quad \text{for} \quad t \in (c - h, c],$$

and analogously for $\mu_{r_2}$. For $\sigma_{r_1}$ we obtain $\sigma_{r_1}^2(h,t) = \sigma_1^2$ for $t \leq c - h$, $\sigma_{r_1}^2(h,t) = \sigma_2^2$ for $t > c$ and

$$\sigma_{r_1}^2(h,t) = \frac{\mu_1\mu_2(t + h - c)(c - t)[(\sigma_1 - \sigma_2)^2 + (\mu_1 + \mu_2)^2] + [(t + h - c)\mu_1\sigma_2 + (c - t)\mu_2\sigma_1]^2}{[(c - t)\mu_2 + (t + h - c)\mu_1]^2}$$

for $t \in (c - h, c]$, and analogously for $\sigma_{r_2}^2$. For a proof and details see Messer (2014), pages 68, Proposition 4.2.4 and Corollary 4.2.9.

Under the null hypothesis we find $\Delta = 1$, such that convergence (26) reduces to (25). Under the alternative of a change point, the estimation of $s$ is not consistent in the range $|t-c| < h$. In that range, one of the windows overlaps the change point, such that the empirical mean and variance are no longer unbiased estimates of the true parameters. However, due to the same argument, the estimation is correct at the change point $t = c$ where the left window refers to $\Phi_1$ and the right to $\Phi_2$.

Considering the distortion term for applications in which the process parameters need to be estimated, we find the following convergence of the filtered derivative process $G^{(n)}$.

**Proposition 4.2.** Let $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 > 0$, with $\mu_1 \neq \mu_2$. Let $\Phi_1(\mu_1, \sigma_1^2)$ and $\Phi_2(\mu_2, \sigma_2^2)$ be independent RP2s and $c \in \tau_h$ be a change point, so that the sequence $(\hat{\Delta}^{(n)})_{n \geq 1}$ results from $\Phi_1$ and $\Phi_2$ according to Construction (2.7). Then, for $G^{(n)}$, $L$ and $\Delta$ as defined in (27), (9) and (26), we have in $(D[h,T-h], d_{SK})$ as $n \to \infty$

$$G^{(n)} - \Delta \hat{\Delta}^{(n)} \overset{d}{\to} \Delta \cdot L.$$

**Proof of Proposition 4.2.** Since $\hat{\Delta}^{(n)} = m^{(n)}/\hat{s}^{(n)}$, the claim follows directly from

$$\Gamma_t^{(n)} = \left( G_t^{(n)} - \frac{m_t^{(n)}}{\hat{s}_t^{(n)}} \right) \frac{\hat{s}_t^{(n)}}{\hat{s}_t^{(n)}}$$

and due to the weak convergence $\Gamma_t^{(n)} \to L$ as stated in Proposition 3.4 and the almost sure convergence $s^{(n)}/\hat{s}^{(n)} \to \Delta$ as in (26) by applying Slutsky’s theorem.

5 Summary

We have extended a convergence result of a filtered derivative process described by Steinebach and Eastwood (1995) and Messer et al. (2014) that can be used for change point analysis in point processes. While usually, the behavior of the filtered derivative process is only described under the null.
Figure 4: Two examples of the function $G_t$ and its connection to the shark function $\Lambda_t$ and the distortion $\Delta_t$. The underlying point process on $[0,1000]$ starts in a process with i.i.d. $\Gamma(p_1, \lambda_1)$-distributed life times and jumps into a process with i.i.d. $\Gamma(p_2, \lambda_2)$-distributed life times at time $c = 500$. Panels A,D: The distortion $\Delta_t$. B,E: The distorted shark function $\Delta_t\Lambda_t$ (solid), the undistorted shark function (dotted, thick) and the hat function (dotted, thin). C,F: The process $G_t$ (solid) that fluctuates around the distorted shark function (dotted). For panels A-C, $(p_1, \lambda_1) = (1, 5), (p_2, \lambda_2) = (1/4, 5)$, resulting in $(\mu_1, \sigma_1^2) = (1/5, 1/25)$ and $(\mu_2, \sigma_2^2) = (1/20, 1/100)$. For panels D-F, $(p_1, \lambda_1) = (2, 10), (p_2, \lambda_2) = (2, 20)$, resulting in $(\mu_1, \sigma_1^2) = (1/5, 1/50)$ and $(\mu_2, \sigma_2^2) = (1/20, 1/200)$. The window size was $h = 150$.

In summary, Proposition 4.2 states that one can approximate (roughly)

$$G^{(n)} \sim \Delta \cdot \left( \Lambda^{(n)} + L \right).$$

The first part, $\Delta \cdot \Lambda^{(n)}$ describes the deterministic, distorted shark function (see Figure 4 for examples), where the distortion is caused by the need to estimate the process parameters and the shark function is a systematically deformed hat whose deformation originates from the scaling that is required for process convergence. The second part, $\Delta \cdot L^{(n)}$, describes a random fluctuation with zero expectation and variance given as the squared distortion.
Acknowledgements

This work was funded by the LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF) and by the Priority Program 1665 of the DFG.

References

Antoch, J. and Hušková, M. (1994). Procedures for the detection of multiple changes in series of independent observations. In Asymptotic statistics (Prague, 1993), Contrib. Statist., pages 3–20. Physica, Heidelberg.

Antoch, J., Hušková, M., and Prášková, Z. (1997). Effect of dependence on statistics for determination of change. J. Statist. Plann. Inference, 60(2):291–310.

Basseville, M. and Nikiforov, I. V. (1993). Detection of abrupt changes: theory and application. Prentice Hall Information and System Sciences Series. Prentice Hall Inc., Englewood Cliffs, NJ.

Bertrand, P. (2000). A local method for estimating change points: the “hat-function”. Statistics, 34(3):215–235.

Bertrand, P. R., Fhima, M., and Guillin, A. (2011). Off-line detection of multiple change points by the filtered derivative with p-value method. Sequential Anal., 30(2):172–207.

Billingsley, P. (1999). Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Brodsky, B. E. and Darkhovsky, B. S. (1993). Nonparametric methods in change-point problems, volume 243 of Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht.

Carlstein, E. (1988). Nonparametric change-point estimation. Ann. Statist., 16(1):188–197.

Cox, D. R. (1962). Renewal theory. Methuen & Co. Ltd., London.

Cox, D. R. and Isham, V. (1980). Point processes. Chapman & Hall, London. Monographs on Applied Probability and Statistics.

Csörgő, M. and Horváth, L. (1987). Asymptotic distributions of pontograms. Math. Proc. Cambridge Philos. Soc., 101(1):131–139.

Csörgő, M. and Horváth, L. (1997). Limit theorems in change-point analysis. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester. With a foreword by David Kendall.

Daley, D. J. and Vere-Jones, D. (2008). An introduction to the theory of point processes. Vol. II. Probability and its Applications (New York). Springer, New York, second edition. General theory and structure.

Davis, R. A., Huang, D. W., and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. Ann. Statist., 23(1):282–304.
Dümbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.*, 19(3):1471–1495.

Gombay, E. and Horváth, L. (1990). Asymptotic distributions of maximum likelihood tests for change in the mean. *Biometrika*, 77(2):411–414.

Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.*, 21(2):671–680.

Horváth, L. (1995). Detecting changes in linear regressions. *Statistics*, 26(3):189–208.

Horváth, L., Kokoszka, P., and Steinebach, J. (2007). On sequential detection of parameter changes in linear regression. *Statist. Probab. Lett.*, 77(9):885–895.

Hušková, M. and Kirch, C. (2012). Bootstrapping sequential change-point tests for linear regression. *Metrika*, 75(5):673–708.

Hušková, M., Kirch, C., Prášková, Z., and Steinebach, J. (2008). On the detection of changes in autoregressive time series. II. Resampling procedures. *J. Statist. Plann. Inference*, 138(6):1697–1721.

Hušková, M., Prášková, Z., and Steinebach, J. (2007). On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist. Plann. Inference*, 137(4):1243–1259.

Hušková, M. and Slabý, A. (2001). Permutation tests for multiple changes. *Kybernetika (Prague)*, 37(5):605–622.

Kendall, D. G. and Kendall, W. S. (1980). Alignments in two-dimensional random sets of points. *Adv. in Appl. Probab.*, 12(2):380–424.

Messer, M. (2014). *A multiple filter test for the detection of rate changes in renewal processes with varying variance*. PhD thesis, Frankfurt, Goethe University, Diss., 2014.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., and Schneider, G. (2014). A multiple filter test for the detection of rate changes in renewal processes with varying variance. *Ann. Appl. Statist.*, to appear. arXiv:1303.3594 [stat.AP].

Muhsal, B. C. S. (2013). *Change-point methods for multivariate autoregressive models and multiple structural breaks in the mean*. PhD thesis, Karlsruhe, Karlsruher Institut für Technologie (KIT), Diss., 2013.

Steinebach, J. and Eastwood, V. R. (1995). On extreme value asymptotics for increments of renewal processes. *J. Statist. Plann. Inference*, 45(1-2):301–312. Extreme value theory and applications (Villeneuve d’Ascq, 1992).

Steinebach, J. and Zhang, H. Q. (1993). On a weighted embedding for pontograms. *Stochastic Process. Appl.*, 47(2):183–195.

Yao, Y.-C. and Davis, R. A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyā Ser. A*, 48(3):339–353.