Asymptotic Behavior of Confidence Regions in the Change-Point Problem

Michael Baron and Andrew L. Rukhin∗
Department of Mathematics and Statistics
UMBC, Baltimore MD 21228

Abstract. The confidence estimation of the change-point is considered. The asymptotic behavior of the coverage probability and the expected width of a traditional confidence region is derived as the threshold constant increases. The nature of this bound is related to information numbers and first passage probabilities for random walks. In the situation when pre- and after-change distributions belong to one-parameter exponential family but are otherwise unknown, similar results are obtained for the confidence region based on the maximum likelihood procedure.


Key words and phrases: change-point problem, confidence estimation, credible region, exponential family, information numbers, random walk.

1 Introduction

Let $F$ and $G$ be two different probability distributions with densities $f$ and $g$ and assume that the observed data $(X_1, \ldots, X_\nu, X_{\nu+1}, \ldots, X_n)$ consists of two independent

∗Research supported by NSF grant #DMS-9124009 and NSA Grant MD904-93-H-3015.
parts, the first, \(X_1, \ldots, X_\nu\), being a random sample from the distribution \(F\), and the second random sample, \(X_{\nu+1}, \ldots, X_n\), coming from the distribution \(G\). In other terms \(\nu\) is the change-point, the parameter of interest, whose estimation was initiated by Chernoff and Zacks (1966), Hinkley (1970), Cobb (1978). The confidence estimation of the change-point was considered by Worsley (1986) and Siegmund (1988).

We study the decision theoretic aspect of this problem assuming the loss function of the form

\[
W(\nu, C) = \lambda|C| + 1 - I_C(\nu). \tag{1}
\]

Here \(C\) is the confidence set with \(|C|\) elements; \(I\) denotes the indicator function and \(\lambda\) is a fixed positive constant. Thus the corresponding risk, \(\lambda E|C| + P(\nu \notin C)\), is a combination of the expected cardinality of the confidence region and of the probability of non-coverage. This risk appears also in the minimization problem of the expected width of a confidence interval with a fixed confidence coefficient.

If distributions \(F\) and \(G\) are known, the classical credible confidence region for \(\nu\) has the form

\[
R_c = \left\{ k : \max_k \left[ \sum_{j=1}^k \log f(X_j) + \sum_{j=k+1}^n \log g(X_j) \right] < \left[ \sum_{j=1}^k \log f(X_j) + \sum_{j=k+1}^n \log g(X_j) \right] + c \right\}
\]

for some positive \(c\).

As is shown in Section 2, the risk function corresponding to this set possesses the following property: as \(\lambda \to 0\), then in order to minimize this risk, the threshold \(c\) must go to infinity. For this reason we look at the asymptotic behavior when \(c \to \infty\) for arbitrary pre-change and after-change distributions. Thus we follow the setting of Siegmund (1988) who derived asymptotic results for large \(c\) for normal distributions \(F\) and \(G\). From the statistical point of view, small values of \(\lambda\) correspond to the dominating role of the coverage probability.

The paper is organized as follows. Results concerning the asymptotic behavior of the expected width, \(E|R_c|\), and of the coverage probability, \(P(\nu \in R_c)\), for \(c \to \infty\) are stated in Section 2. In Section 3 we study the situation when densities \(f\) and \(g\) are not completely known, but are assumed to belong to the same one-parameter exponential family. It is shown that the confidence region based on the likelihood ratio test for any fixed (but unknown) values of the nuisance parameters exhibit the same asymptotic behavior when these parameters are known (and this behavior is described in Section 2). Proofs of all these results are collected in Section 4.
2 Asymptotic behavior of the credible confidence region

To study the confidence estimation problem described in Section 1 we consider the following asymptotic setting originally suggested by Cobb (1978). Let \( \{X_j, j \in \mathbb{Z}\} \) be a doubly infinite sequence of independent random variables. Assume that for some integer \( \nu \) \( X_j \) has density \( f(x) \) if \( j \leq \nu \) and density \( g(x) \) otherwise. Here we assume that \( f \) and \( g \) are known positive densities.

To relate this problem to the situation described in Section 1 with \( m = \left[\frac{n-1}{2}\right] \) relabel the original observations as \( X_{-m}, X_{-m+1}, \ldots, X_1, X_0, X_1, \ldots, X_{n/2} \). Then as \( n \to \infty \), the setting above arises. For practical applications the obtained confidence regions have to be intersected with the set \(-m \leq \nu \leq \lfloor n/2 \rfloor\).

Let \( z_j = \log \frac{f}{g}(X_j) \) and for \( k \geq 1 \) \( S_k = \sum_{j=1}^{k} z_j \), \( S_k = -\sum_{j=k+1}^{0} z_j \) if \( k \leq -1 \); \( S_0 = 0 \). Then the credible confidence set introduced in Section 1 has the form

\[
R_c = \{k : S_k < \max_k S_k - c\}
\]

for some positive constant \( c \).

For \( j > \nu \), \( E z_j = -\rho_1 < 0 \), and \( E z_j = \rho_2 > 0 \) for \( j \leq \nu \), with \( \rho_1, \rho_2 \) denoting the information numbers, which are assumed to be finite. According to the strong law of large numbers \( \max_k S_k \) is finite with probability one and \( R_c \) is well defined.

Throughout this paper, if \( S_1 \) is an arithmetic random variable with a span \( d \), i.e. its support is contained in \( \{..., -2d, -d, 0, d, 2d, ...\} \) with \( d \) being the largest constant with this property, \( c \) will tend to \( +\infty \) only through multiples of \( d \).

Theorem 2.1 For \( c \to \infty \)

\[
1 - P\{\nu \in R_c\} \sim e^{-c\pi_1\pi_2} \left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right),
\]

where

\[
\pi_1 = \exp \left\{ -\sum_{k=1}^{\infty} k^{-1} P_G(S_k > 0) \right\}, \quad \pi_2 = \exp \left\{ \sum_{k=-\infty}^{-1} k^{-1} P_F(S_k \geq 0) \right\},
\]

and \( \rho_1, \rho_2 \) are information numbers.

A similar result for point estimation of the change-point parameter for the zero-one loss function is given in Rukhin (1995).

The next result pertains to the asymptotics of the expected width.
Theorem 2.2 Let $F$ and $G$ satisfy to one of the following conditions:

(i) $FG$ are continuous distributions.

(ii) $FG$ are mixed type distributions (i.e. they have both discrete and continuous components), and

$$\sum^* e^x P\{S_{\tau^+} = x\} < 1,$$

where $\sum^*$ means the summation over only discrete values of $S_{\tau^+}$.

Then the average length of $R_c$ for $c \to \infty$ has the form

$$E|R_c| = c \left[\frac{1}{\rho_1} + \frac{1}{\rho_2}\right] + b + o(1),$$

where $b$ is a constant term specified in (26).

According to Theorems 2.1 and 2.2, the risk of the confidence set $R_c$ for the loss function (1) has the asymptotic representation as $c \to \infty$

$$\text{Risk}(W, R_c) = (\lambda c + e^{-c\pi_1\pi_2}) \left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) + \lambda(b + o(1)).$$

Constant $c$ which minimizes this risk, admits the following asymptotic representation for $\lambda \to 0$

$$c \sim -\log \lambda + \log(\pi_1\pi_2).$$

With this choice of $c$,

$$\text{Risk}(W, R_{c(\lambda)}) = -\lambda \log \lambda \left[\frac{1}{\rho_1} + \frac{1}{\rho_2}\right] + \lambda \left[\frac{(1 + \log(\pi_1\pi_2))(\rho_1 + \rho_2)}{\rho_1\rho_2}\right] + b + o(\lambda)$$

as $\lambda \to 0$.

If $F = N(0, 1)$ and $G = N(\Delta, 1)$, then $\rho_1 = \rho_2 = \Delta^2/2$. According to Theorems 2.1 and 2.2 as $c \to \infty$

$$E|R_c| \sim \frac{4c}{\Delta^2}, \quad 1 - P_{\nu\{\nu \in R_c\}} \sim \frac{4\exp(-c - \kappa)}{\Delta^2},$$

where $\kappa = 2 \sum_1^\infty k^{-1} \Phi(-2\Delta \sqrt{k})$. This agrees with results of Siegmund (1988).

As another example consider the situation when $F$ and $G$ are Bernoulli distributions. Then, by a translation invariant property of $P_{\nu\{\nu \in R_c\}}$, $P_{\nu\{\nu \in R_c\}} = P_{\nu\{0 \in R_c\}} = P_{\nu\{\max_{k \leq 0} S_k < c\}} P_{G\{\max_{k \geq 0} S_k < c\}} = \tilde{Q}^2(c).$
Clearly, $\tilde{Q}(c)$, the left-continuous version of the distribution function of $\max_{k \leq 0} S_k$, satisfies the following integral equation

$$\tilde{Q}(c) = \int P_F\{z_1 \in du\} \tilde{Q}(c-u) = E_F \tilde{Q}(c-z_1) \text{ for } c \geq 0. \quad (7)$$

Also it vanishes for non-positive $c$ and monotonically increases to 1 as $c$ increases to infinity.

The solution to (7), when probabilities of success for $F$ and $G$ are $p$, $0 < p < 1/2$, and $(1-p)$ respectively, turns out not to depend on the value of $p$. Indeed $z_1$ takes values $\zeta$ and $-\zeta$ with probabilities $p$ and $1-p$ where $\zeta = \log \frac{1-p}{p}$ and $\max_{k \leq 0} S_k$ can take only the values which are multiples of $\zeta$. Hence the function $\tilde{Q}(c)$ is a constant on any interval $(n\zeta, (n+1)\zeta]$ with integer $n$. Put $c = n\zeta$ and $a_n = \tilde{Q}(c)$. Then $a_n \uparrow 1$ as $n \rightarrow \infty$, $a_n = 0$ for $n \leq 0$, and (7) implies the following difference equation for $a_n$:

$$a_n = pa_{n-1} + (1-p)a_{n+1}. \quad (8)$$

The unique solution has the form $a_n = 1 - \exp \{-n\zeta\}, \ n > 0$, and

$$\tilde{Q}(c) = a_c/\zeta = 1 - e^{-c}. \quad (8)$$

Therefore,

$$P_{\nu}\{\nu \in R_c\} = (1 - e^{-c})^2, \quad (9)$$

which, rather surprisingly, does not depend on $p$. Thus in this example the set $R_c$ has the guaranteed confidence coefficient for all values of $p$.

### 3 Confidence estimation of the change-point when $F$ and $G$ are unknown

Suppose now that distributions $F$ and $G$ belong to an exponential family but otherwise are unknown. More precisely assume that both $f$ and $g$ have the form

$$f(x|\theta) = \exp\{\theta x - \psi(\theta)\} f(x|0)$$

with some real $\theta_1$ and $\theta_2$, $\theta_2 \neq \theta_1$ (although our results can be extended to the multivariate case). Function $\psi(\theta)$ is known to be an analytic convex function on $\Theta$, which is a subinterval of the real line (see for example Brown, 1986). We assume that interval $[\theta_1, \theta_2]$ is contained in $\text{Int} \ \Theta$.

It is also assumed that for some $0 < \omega \leq 1$ and $l > 0$

$$\min\{\nu, n - \nu\} \geq ln^\omega. \quad (10)$$
Thus our problem of interval estimation of \( \nu \) can be considered as the Bayes estimation problem with the discrete uniform prior on the set defined by (10), and the risk function \( \lambda E|R| + P\{\nu \not\in R\} \), when \( \theta_1 \) and \( \theta_2 \) are nuisance parameters.

For any fixed \( 1 \leq k \leq n \) let \( \hat{\theta}_1 = \hat{\theta}_1(k) \) and \( \hat{\theta}_2 = \hat{\theta}_2(k) \) be the maximum likelihood estimators for the parameters based on \( X_1, \ldots, X_k \) and \( X_{k+1}, \ldots, X_n \) respectively. Then the Bayes estimator of \( \nu \) with respect to the zero-one loss and the given prior is

\[
\hat{\nu} = \arg \max \{ \Lambda_k, \ln \omega \leq k \leq n - \ln \omega \},
\]

where

\[
\Lambda_k = \log \prod_{j=1}^{k} f(X_j|\hat{\theta}_1) \prod_{j=k+1}^{n} f(X_j|\hat{\theta}_2).
\]

Define for a positive \( c \)

\[
T_c = \{ k : \ln \omega \leq k \leq n - \ln \omega, \Lambda_\nu - \Lambda_k < c \}.
\]

Our goal is to study the behavior of \( T_c \) in terms of the given risk function and to obtain the optimal value of \( c \). The formulas for the asymptotic risk of \( T_c \) as \( n \to \infty \) follow from the proximity of the set \( T_c \) and of the set \( R_c \), defined by (2) for known \( \theta_1 \) and \( \theta_2 \).

**Proposition 3.1** One has \( P\{T_c \subset [\nu - n^{\alpha}, \nu + n^{\alpha}]\} \to 1 \) as \( n \to \infty \) for any \( 0 < \alpha < \omega \) and

\[
1 - P\{T_c \subset [\nu - n^{\alpha}, \nu + n^{\alpha}]\} = o\left(e^{-Kn^{\alpha}}\right)
\]
as \( n \to \infty \) for some positive constant \( K \).

**Proposition 3.2** Let \( \varepsilon_n = r n^{-\beta} \) for some \( 0 \leq \beta < \omega/2 - \alpha \) and \( r > 0 \). Then

\[
P\left\{ \max_{[\nu - n^{\alpha}, \nu + n^{\alpha}]} \left| (\Lambda_\nu - \Lambda_k) - (S_\nu - S_k) \right| \geq \varepsilon_n \right\} = o\left(\exp\left\{-Cn^{(\omega - 2\alpha - 2\beta)/3}\right\}\right)
\]
as \( n \to \infty \) for some \( C > 0 \).

As was observed, \( R_c = \{S_\nu - S_k \leq c\} \subset [\nu - n^{\alpha}, \nu + n^{\alpha}] \) with probability tending to one. Therefore Propositions 3.1 and 3.2 imply the first part of the following result.

**Theorem 3.3** Let \( \varepsilon_n = r n^{-\beta} \) for some \( 0 \leq \beta < \omega/2 \) and \( r > 0 \). Then

\[
P\{R_{c-\varepsilon_n} \subset T_c \subset R_{c+\varepsilon_n}\} \to 1 \text{ as } n \to \infty.
\]
The rate of convergence is \( o\left(\exp\left\{-Kn^{(\omega - 2\beta)/5}\right\}\right) \) for some \( K > 0 \). Here \( c \) can be any bounded positive function of \( n \).
Theorem 3.3, whose proof along with the proofs of Propositions is given in Appendix, states that the sets $T_c$ and $R_c$ are very close when $n$ is large, so that

$$\lim_{n \to \infty} P\{\nu \in R_{c - \epsilon_n}\} \leq \lim_{n \to \infty} P\{\nu \in T_c\} \leq \lim_{n \to \infty} P\{\nu \in R_{c + \epsilon_n}\}$$

and

$$\lim_{n \to \infty} E|R_{c - \epsilon_n}| \leq \lim_{n \to \infty} E|T_c| \leq \lim_{n \to \infty} E|R_{c + \epsilon_n}|.$$}

Letting $n \to \infty$ in both inequalities and using Theorems 2.1 and 3.1 one derives the following corollaries.

**Corollary 1** As $c \to +\infty$

$$\lim_{n \to \infty} (1 - P\{\nu \in T_c\}) \sim e^{-c \pi_1 \pi_2} \left( \frac{1}{\rho_1} + \frac{1}{\rho_2} \right),$$

where $\pi_1$ and $\pi_2$ are given by (4).

**Corollary 2** As $c \to +\infty$

$$\lim_{n \to \infty} E|T_c| \sim c \left( \frac{1}{\rho_1} + \frac{1}{\rho_2} \right).$$

Thus the risk function of $T_c$ behaves asymptotically like that of $R_c$. Therefore the best choice of $c$ in terms of the Bayes risk when $\lambda \to 0$ is

$$c \sim -\log \lambda + \log(\pi_2 \pi_2).$$

**Corollary 3** If distributions $F$ and $G$ do not have atoms, then

$$\lim_{n \to \infty} P\{\nu = \hat{\nu}\} = P\{\tau_+ = \infty\} P\{\tau'_+ = \infty\},$$

(13)

where $\tau_+ = \inf\{k > 0; S_{\nu + k} \geq 0\}$ and $\tau'_+ = \inf\{k > 0; S_{\nu - k} \geq 0\}$.

To prove (13) note that because of the absence of atoms $\nu = \hat{\nu}$ is $\mu$-almost surely equivalent to $\Lambda_\nu = \Lambda_{\hat{\nu}}$ or to the event $\nu \in T_0$. Let $\epsilon_n = n^{-\beta}$ for some $0 \leq \beta < \omega/2$ so that $T_0 = \bigcap T_{\epsilon_n}$. According to Theorem 3.3 $R_0 \subset T_{\epsilon_n} \subset R_{2\epsilon_n}$ with probability tending to one. Hence as $n \to \infty$ the limiting probability $P\{\nu = \hat{\nu}\}$ is bounded from below by $P\{\nu \in R_0\} = P\{S_{\max} = S_{\nu}\} = P\{\tau_+ = \infty\} P\{\tau'_+ = \infty\}$ and from above by $\lim_{n \to \infty} P\{\nu \in R_{2\epsilon_n}\} = P\{\nu \in R_0\}$. Thus (13) holds.

Notice that under condition (10) one can also estimate $\theta_1$ on the basis of the data $X_i, 1 \leq i < ln^\omega$ and $\theta_2$ on the basis of the data $X_j, n - ln^\omega < j \leq n$. It is possible to show that the truncated random walk with so estimated parameters converges to the random walk with known $\theta_1$ and $\theta_2$. In this way another version of asymptotically optimal confidence region can be obtained. However, the confidence set $T_c$ seems to be more robust to the choice of $l$ and $\omega$. 8
4 Appendix

In our setting the quantities \( P \{ \nu \in R_c \} \) and \( E|R_c| \) do not depend on \( \nu \), so we can assume that \( \nu = 0 \). Then \( \{ S_k, k \in \mathbb{Z} \} \) can be viewed as a combination of two independent random walks: \( S_k^{(1)} = S_{-k} \) and \( S_k^{(2)} = S_k, k \geq 0 \). We attach index \( j \) to all probabilities and expected values related to \( S_k^{(j)} \), \( j = 1, 2 \).

Let \( M = \max_{-\infty < k < \infty} S_k; M_j = \max_{0 \leq k < \infty} S_k^{(j)} \) for \( j = 1, 2 \) and denote by \( Q(u), Q_1(u) \) and \( Q_2(u) \) distributions of \( M, M_1 \) and \( M_2 \) respectively. Also for \( j = 1, 2 \) we introduce the first passage moments,

\[
\tau^{(j)}(x) = \begin{cases} 
\inf \{ k : S_k^{(j)} \geq x \} & \text{for } x > 0; \\
\inf \{ k : S_k^{(j)} \leq x \} & \text{for } x < 0; 
\end{cases}
\]

\[
\tau^+_j = \inf \{ k : S_k^{(j)} > 0 \}; \quad \tau^-_j = \inf \{ k : S_k^{(j)} < 0 \}.
\]

4.1 Proof of Theorem 2.1

We can think of \( f \) and \( g \) as imbedded into one-parameter exponential family of the form

\[
f(x|\theta) = \exp \left\{ \theta \log \frac{f(x)}{g(x)} - \psi(\theta) \right\} g(x).
\]

Then \( \exp \{ \psi(\theta) \} = \int f^\theta g^{1-\theta} \), \( f(x|1) = f(x) \) and \( f(x|0) = g(x) \). Moreover, \( \psi'(0) = -\rho_2 \) and \( \psi'(1) = \rho_1 \). Since

\[
P \{ \nu \in R_c \} = P \{ M_1 < c \} P \{ M_2 < c \} = (1 - P_1 \{ \tau(c) < \infty \})(1 - P_2 \{ \tau(c) < \infty \}) \tag{14}
\]

we can use formula (8.48) of Siegmund (1985) which shows that for \( c \to \infty \)

\[
P_1 \{ \tau(c) < \infty \} \sim e^{-c} P_1 \{ \tau_+ = \infty \} P_2 \{ \tau_- = \infty \} / \psi'(1) = e^{-c} \pi_1 \pi_2 / \rho_1. \tag{15}
\]

Similar asymptotics obtains for \( P_2 \{ \tau(c) < \infty \} \), so that Theorem 2.1 follows from (14).

4.2 Proof of Theorem 2.2

In our notation

\[
E|R_c| = \int_{u \geq 0} E \{|R_c| \mid M = u\} \ dQ(u) = \int_{u \geq 0} E \{|R_c| \mid M_1 = u; \ M_2 \leq u\} \ Q_2(u) \ dQ_1(u)
\]
\[ + \int_{u \geq 0} E \{ |R_c| \mid M_1 = u; M_2 = u \} Q_1(u^-) \, dQ_2(u), \quad (16) \]

where \(Q_1(u^-)\) denotes the limit from the left of \(Q_1(u_1)\) as \(u_1 \to u\). We consider only the first integral in the right-hand side of (16), the second integral can be treated similarly.

The set \(R_c\) may be disconnected (i.e. it can have lacunas). Denote connected components of \(R_c\) to the left of the origin by \(G_0, G_1, \ldots\), starting from the origin, and the components to the right of it by \(H_0, H_1, \ldots\). Here we put \(G_0 = \emptyset\) if \(0 \notin R_c\), and \(H_0 = \emptyset\) if \(1 \notin R_c\).

Then

\[ |R_c| = \sum_{i=0}^{n_1} |G_i| + \sum_{i=0}^{n_2} |H_i|, \]

where \(n_1\) and \(n_2\) are the (random) numbers of connected components.

One has for any positive \(u\)

\[ E \{ |R_c| \mid M_1 = u; M_2 \leq u \} \quad (17) \]

Considering separately the four terms that appear in (17), let

\[ G_{u,c} = \{ k \leq \tau(u-c) : S_k^{(1)} > u-c \}, \quad \text{if} \quad u \leq c; = \emptyset, \quad \text{otherwise.} \]

Then, under the condition \(M_1 = u\), one has \(G_0 = G_{u,c}\), and

\[ E \{ |G_0| \mid M_1 = u \geq M_2 \} = E \{ |G_{u,c}| \mid M_1 = u \} \]

\[ = E \{ |G_{u,c}| \mid \max_{G_{u,c}} S_k = u \} \, P_1 \{ \max_{G_0} S_k = M_1 \mid M_1 = u \} \]

\[ + E \{ |G_{u,c}| \mid \max_{G_{u,c}} S_k < u \} \, P_1 \{ \max_{G_0} S_k < M_1 \mid M_1 = u \}, \quad (18) \]

where for \(u < c\)

\[ E_1 \{ |G_{u,c}| \mid \max_{G_{u,c}} S_k = u \} \]

\[ = E_1 \{ \tau(u) \mid \tau(u) < \tau(u-c), S_{\tau(u)} = u \} + E_1 \{ \tau(-c) \mid \tau_+ = \infty \}. \quad (19) \]

**Lemma 1** Under the hypothesis of Theorem 2.2,

\[ E_1 \{ \tau(-c) \mid \tau_+ = \infty \} = \rho_1^{-1} \left( c + \frac{E_1 Z_1^2}{2\rho_1} - 2E_1 M_1 \right) - E_1 \tau(M_1) + o(1), \quad (20) \]

as \(c \to \infty\).
Proof: Consider the queueing process $W_k$, defined as $W_0 = 0; W_{k+1} = \min\{0; W_k - Z_{k+1}\}$ for $k \geq 0$.

Observe that given $\tau_+ = \infty$, both processes $W_k S_{k}^{(1)}$ take only non-positive values. Therefore, under this condition they coincide for all $k$. Hence, for $\tau^W(-c) = \inf\{k : W_k \leq -c\}$,

$$E_1\{\tau(-c)|\tau_+ = \infty\} = E_1\{\tau^W(-c)|\tau_+ = \infty\}. \quad (21)$$

It is easy to see that $W_k = S_{k}^{(1)} - \max_{j \leq k} S_{j}^{(1)}$. Then $W_{\tau(M_1)} = 0$, and the process $W_k^* = W_{k + \tau(M_1)} - W_{\tau(M_1)}$ is distributed as $W_k$ conditioned on $\tau_+ = \infty$. Then, since $\tau^W(-c) = \tau(M_1) + \tau^W^*(-c)$, it follows that

$$E_1\{\tau^W(-c)|\tau_+ = \infty\} = E_1\tau^W(-c) = E_1\tau^W(-c) - E_1\tau(M_1).$$

The following result of Lotov (1991) can be used to evaluate the asymptotics of $E_1\tau^W(-c)$.

**Theorem** (Lotov, 1991) Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with $E\xi_1 > 0$ and $f(\lambda) = E e^{-\lambda \xi_1}$. Let $S_k = \xi_1 + \ldots + \xi_k$ and let $Y_k$ be the corresponding queueing process, $Y_0 = 0$, $Y_{n+1} = \max(0, Y_n + \xi_{n+1})$. Assume that

a) $|f(\lambda)| < \infty$ for $-\gamma \leq \Re \lambda \leq \beta$ ($\gamma \geq 0$, $\beta > 0$).

b) $f(\beta) \geq 1$; if $f(\beta) = 1$, then also $\int_0^\infty ye^{\beta y} P\{-\xi_1 \leq dy\} < \infty$.

c) $\tau(\lambda) = 1 - f(\lambda)$ for $-\gamma \leq \Re \lambda \leq \beta$ has only two zeros, $\lambda = 0$ and a unique positive root $\lambda = \eta$.

Let $\eta_\pm = \inf\{n \geq 1 : S_n \geq 0\}$, (inf $\emptyset$ = $\infty$), $\chi_\pm = S_{\eta_\pm}$, $G_1(y) = P\{\chi_\pm < y, \eta_\pm < \infty\}$, $G_2(y) = P\{-\chi_\pm < y\}$. Also suppose that

da) $\int_0^\infty e^{\beta y} dG_1^0(y) < 1$, $\int_0^\infty e^{\gamma y} dG_2^0(y) < 1$, where $G_j^0$ denotes the sum of the discrete and singular components of $G_j$ ($j = 1, 2$).

Then for the stopping time $T = T(b) = \min\{n \geq 1 : Y_n > b\}$,

$$ET = (E\xi_1)^{-1}\left\{b + \frac{E\xi_1^2}{2E\xi_1} \frac{2r_+^2(0)}{r_+(0)} + \frac{q_1 q_2 e^{-\phi b}}{q} + o(e^{-\gamma b}) + o(e^{-\beta b})\right\},$$

as $b \to \infty$, where $r_+^\pm(\lambda) = 1 - E(\exp\{\lambda \chi_\pm\} - e^{\gamma \chi_\pm})$, $q_1 = -r_+(0)/qr_+(q)$, $q_2 = r_-(q)/qr_-(0)$.

We let $\xi_j = z_j b = c$. Then, in our notation, $W_k = -Y_k$, $\tau^W(-c) = T(b)$, $f(\lambda) = \int f^{1-\gamma^\lambda} \gamma^\lambda E\xi_1 = \rho_1$, $\chi_\pm = S_{\tau_\pm}^{(1)}$, $r_+(0) = 1 - P\{\tau_+ < \infty\} = \eta_1$, and $r_+(0) = -E(S_{\tau_+} + \tau_+ < \infty)$. Conditions (a) and (c) hold for $\gamma = 0, \beta = 1$. Also, for $\beta = 1$, $\int_0^\infty ye^{\beta y} P\{-\xi_1 < dy\} \leq E(-z_{-1}) e^{-z_{-1}} = E_G \log \frac{q}{f}(X) = \rho_2 < \infty$, so that
(b) holds. Clearly, with \( \gamma = 0 \),
\[
\int_0^\infty e^{\gamma y} dG_0(y) < 1
\]
for continuous and mixed type distributions. Inequality
\[
\int_0^\infty e^{\gamma y} dG_1(y) < 1
\]
is trivial for continuous distributions, and it is guaranteed by (5) for mixed type \( F \) and \( G \). Thus, (d) holds.

Hence, all the conditions of Lotov’s Theorem are met, and therefore,
\[
E_1 \tau^W(-c) = \rho^{-1}(c + \beta_1) + o(1),
\]
as \( c \to \infty \), where \( \beta_1 = E_1 Z_1^2 / 2\rho_1 - 2E_1 \{ S_{\tau+} | \tau < \infty \} / \eta_1 \).

Next, we note that \( M_1 = \sum_{k=1}^{q} T_k \), where \( T_1, \ldots, T_q \) are consecutive ascending ladder heights of the random walk \( S^{(1)} \). The common distribution of \( T_k \) coincides with the conditional distribution of \( S_{\tau+} \) given \( \tau_+ = \infty \), and \( q \) is an independent geometric random variable with parameter \( P_1 \{ \tau_+ = \infty \} = \eta_1 \). Hence,
\[
E_1 \{ S_{\tau+} | \tau_+ < \infty \} = E T_1 = E_1 M_1 / E q = \eta_1 E_1 M_1,
\]
and (20) follows from (21) and (22).

Our next step is to show that only the first term in (18) matters.

**Lemma 2** For any \( u \) from the support of \( Q_1 \),
\[
-c^{-1} \log P \left\{ \max_{G_0} S_k^{(1)} < M_1 \mid M_1 = u \right\} \geq 1 + o(1),
\]
as \( c \to \infty \).

**Proof:** In terms of the first passage time,
\[
P \{ \max_{G_0} S_k^{(1)} < M_1 \mid M_1 = u \} = P_1 \{ \tau(u - c) < \tau(u) \mid M_1 = u \}
\]
\[
= P_1 \{ \tau(u - c) < \tau(u) \mid \tau(u) < \infty \} = \frac{P_1 \{ \tau(u - c) < \tau(u) < \infty \}}{P_1 \{ \tau(u) < \infty \}}.
\]
Consider the random walk \( S_k^* = \sum_{k=1}^{\tau(u-c)} - S_{\tau(u-c)}^{(1)} \) for \( k > 0 \). For any \( c \) the moment \( \tau(u - c) \) is a Markov stopping time. Hence by the strong Markov property, the process \( S_k^* \) is independent on \( \tau(u-c) \) and has the same distribution as \( S_k^{(1)} \). The
first passage time $\tau^*_c = \inf\{k \geq 1 : S^*_k \geq c\}$, which is a function of $S^*_k$, is identically distributed with $\tau^*_c$. According to (15),

$$P_1\{\tau(u - c) < \tau(u) < \infty\} \leq P_1\{\tau^*_c < \infty\} \sim e^{-c} \pi_1 \pi_2 / \rho_1,$$

as $c \to \infty$, which implies (24).

Using the notation $I\{A\}$ for the indicator of the event $A$, notice that $\max_{G_{u,c}} S^{(1)}_k$ and $|G_{u,c}| = \sum_{k=0}^{\tau(u-c)} I\{S^{(1)}_k > u - c\}$ are positively associated random variables (see Lehmann (1966)). Indeed, they are non-decreasing functions of $S^{(1)}_1, S^{(1)}_2, \ldots$, and therefore are non-decreasing functions of i.i.d. random variables $z_1, z_2, \ldots$. This property implies that by (23),

$$E_1\{|G_{u,c}| \mid \max_{G_{u,c}} S_k < u\} \leq E_1\{|G_{u,c}| \mid \max_{G_{u,c}} S_k = u\} = O(c).$$

The second term of (18) tends to 0, because it is bounded by a product of linearly increasing and exponentially decreasing functions of $c$. Also, according to Lemma 2

$$P\{\max_{G_{0}} S^{(1)}_k = M_1 \mid M_1 = u\} = 1 + O(e^{-c})$$

as $c \to \infty$.

One obtains from (18) and (23)

$$E\{|G_0| \mid M_1 = u \geq M_2\} = E_1\{\tau(u) \mid \tau(u) < \infty, S_{\tau(u)} = u\} + (c + \beta_1) / \rho_1 + E_1\tau(M) + o(1), \quad c \to \infty.$$

We show now that as $c \to \infty$, $E_1\sum_{i=1}^{n_1} \{G_i \mid M_1 = u\}$ has a finite limit, independent of $u$. It follows from (24) that all moments $k$ defined by $S^{(1)}_k = M_1$ do not belong to $\bigcup_{i \geq 1} G_i$ with probability tending to 1. Hence the limiting conditional distribution of $\sum_{i \geq 1} G_i$ given $M_1 = u$ coincides with its limiting unconditional distribution. In fact it is determined by the limiting distribution of $c + S^{(1)}_{\tau(-c)}$, whose existence is discussed in Siegmund (1985), Chapter VIII.

Conditioned on $M_1 = u$, the moment $\tau(M_1 - c)$ is a Markov stopping time. Hence the processes $S^*_k$ and $S^{(1)}_k$ have the same conditional distribution. Also

$$E_1\left\{\sum_{i=1}^{n_1} G_i \mid M_1 = u\right\} = E_1\left\{\#k : S^*_k > c + S^{(1)}_{\tau(-c)} \mid M_1 = u\right\} \to E[G_+].$$
Here $G_+ = \{ k : S_k^{(1)} > y \}$ and $y$ denotes a random variable independent on $S_k$, whose distribution is the limiting distribution of $c + S_{\tau(-c)}$. 

Consider now the last term of (17). As above,

$$E_2 \left\{ \sum_{i=1}^{n_2} |H_i| \mid M_2 \leq u = M_1 \right\} \rightarrow E_2|H_+|,$$

where $|H_+|$ is defined similarly to $G_+$. Thus only the asymptotic distribution of $|H_0|$ depends on $c$. Applying Lemma 1 to $S_k^{(2)}$ and $\tau(-v + u - c)$, one has as $c \rightarrow \infty$,

$$E_2\{|H_0| \mid M_2 = v \leq M_1 = u \}$$

$$= E_2\{\tau(v) \mid \tau(v) < \tau(u - c), S_{\tau(v)} = v\} + E_2\{\tau(-v + u - c) \mid \tau_+ = \infty\} - 1$$

$$= E_2\{\tau(v) \mid \tau(v) < \infty, S_{\tau(v)} = v\} + (c + v - u + \beta_2)/\rho_2 - E_2\tau(M) - 1 + o(1), \quad (25)$$

where $\beta_2$ is defined similarly to $\beta_1$.

Thus (17), (23) and (25) yield

$$E \{|R_c| \mid M_1 = u \geq M_2 = v\} = c/\rho_1 + c/\rho_2 + w_1(u, v) + o(1),$$

where

$$w_1(u, v) = E_1\{\tau(u) \mid \tau(u) < \infty, S_{\tau(u)} = u\} + E_2\{\tau(v) \mid \tau(v) < \infty, S_{\tau(v)} = v\} + \beta_1/\rho_1 - (u - v - \beta_2)/\rho_2 - E_1\tau(M) - E_2\tau(M) + E_1|G_+| + E_2|H_+| - 1.$$ 

A similar formula holds for $E \{|R_c| \mid M_2 = u > M_1 = v\}$, and $w_2(u, v)$ is defined similarly to $w_1(u, v)$.

Next, we note that

(i) $\int_{u \geq 0} \int_{v \leq u} dQ_2(v)dQ_1(u) + \int_{u \geq 0} \int_{v < u} dQ_1(v)dQ_2(u) = \int_{u \geq 0} dQ(u) = 1,$

(ii) $\int_{u \geq 0} \int_{v \leq u} dQ_j(v)dQ_j(u) = E_j\tau(M)$ for $j = 1, 2$,

(iii) $\int_{u \geq 0} \int_{v \leq u} (u - v)dQ_2(v)dQ_1(u) = E(M_1 - M_2)^+,$

(iv) $M_1 + (M_2 - M_1)^+ = M_2 + (M_1 - M_2)^+ = \max\{M_1, M_2\} = M,$

where $x^+ = \max\{0, x\}$. It follows from (16) that

$$E|R_c| = \int_{u \geq 0} \int_{v \leq u} E \{|R_c| \mid M_1 = u \geq M_2 = v\}dQ_2(v)dQ_1(u)$$

$$+ \int_{u \geq 0} \int_{v < u} E \{|R_c| \mid M_2 = u > M_1 = v\}dQ_1(v)dQ_2(u) = c(1/\rho_1 + 1/\rho_2) + b + o(1),$$

14
as $c \to \infty$. Here

$$b = \int_{u \geq 0} \int_{v \leq u} w_1(u, v) \, dQ_2(v) \, dQ_1(u) + \int_{u \geq 0} \left( \int_{v < u} w_2(u, v) \, dQ_1(v) \right) \, dQ_2(u) \tag{26}$$

$$= \frac{E_1 Z_1^2 / 2 \rho_1}{\rho_1} - E(M + M_1) + \frac{E_2 Z_2^2 / 2 \rho_2}{\rho_2} - E(M + M_2) + E_1 \left| G_+ \right| + E_2 \left| H_+ \right| - 1.$$

By using this approach one can obtain a limiting expression for the probability of $R_c$ being a connected set. Indeed,

$$P\{R_c \text{ is connected} \} = P\{R_c = G_0 \cup H_0 \} = P\{G_1 = \emptyset, H_1 = \emptyset \}.$$

The distribution of $\left| G_1 \right|$ and $\left| H_1 \right|$ is completely determined by the “overshoot” $c + S_{\tau(-c)}^{(j)}$, $j = 1, 2$. If, as in the proof of Theorem 2.2, random variables $y_j, j = 1, 2$ have the distribution coinciding with the limiting distribution of $c + S_{\tau(-c)}^{(j)}$, and $F_j$ denotes the distribution of $M_j = \max_k S_k^{(j)}$, then as $c \to \infty$

$$P\{G_1 = \emptyset \} \to P\{\tau(\left| y_1 \right|) = \infty \} = P\{M_1 < \left| y_1 \right|\} = E \left[ F_1(\left| y_1 \right|) \right]$$

and $P\{H_1 = \emptyset \} \to E \left[ F_2(\left| y_2 \right|) \right]$ and by independence

$$P\{R_c \text{ is connected} \} \to E \left[ F_1(\left| y_1 \right|) \right] E \left[ F_2(\left| y_2 \right|) \right].$$

4.3 Proof of Proposition 3.1

Let $H(x) = \sup_{\theta} \{ \theta x - \psi(\theta) \}$ be the Legendre transform of the function $\psi$. It follows from (12) and the definition of $\hat{\theta}_1$ and $\hat{\theta}_2$ that

$$\Lambda_k = k(\hat{\theta}_1 \bar{x}_k - \psi(\hat{\theta}_1)) + (n - k)(\hat{\theta}_2 \bar{x}_{kn} - \psi(\hat{\theta}_2)) = kH(\bar{x}_k) + (n - k)H(\bar{x}_{kn}). \tag{27}$$

Here and later we use the notation

$$\bar{x}_{kj} = \frac{X_{k+1} + \cdots + X_j}{j - k}, \quad \bar{x}_k = \frac{X_1 + \cdots + X_k}{k}.$$

The maximum likelihood estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, which maximize the function

$$\theta x - \psi(\theta) \text{ for } x = \bar{x}_k \text{ and } x = \bar{x}_{kn},$$

are the roots of the equation $\psi'(\theta) = x$ for these values of $x$. Since $\psi'$ is an increasing function of $\theta$, we can introduce $h(x)$, the inverse function of $\psi'(\theta)$. Then $\hat{\theta}_1 = h(\bar{x}_k)$, $\hat{\theta}_2 = h(\bar{x}_{kn})$ and $H(x) = xh(x) - \psi(h(x))$, so that $H'(x) = h(x)$. Also $\theta_1 = h(\mu_1)$ and $\theta_2 = h(\mu_2)$, where $\mu_j = E_{\theta_j}X$.  

15
Lemma 3 Let \( \xi_1, \xi_2, \ldots, \xi_k \) be i.i.d. random variables with density \( f(x|\theta) \) for some \( \theta \in \text{Int } \Theta \) and \( \mu = \mathbb{E} \xi_1 \). Let \( \Delta \) be a closed subset of \( \Theta \) containing \( \theta \) in its interior and

\[
m = \min \left\{ \frac{1}{\psi''(t)}, \ t \in \Delta \right\} > 0.
\]

Then for any \( \varepsilon > 0 \) such that \( h(\mu - \varepsilon, \mu + \varepsilon]) \subset \Delta \)

\[
P \{ |\tilde{\xi} - \mu| \geq \varepsilon \} \leq 2e^{-ms^2k/2}.
\] (28)

Proof: Since \( \mathbb{E}e^{\lambda \xi} = \exp \{ \psi(\theta + \lambda) - \psi(\theta) \} \), the classical results from the theory of large deviations (see e.g. Bahadur, 1971) imply that

\[
P \{ |\tilde{\xi} - \mu| \geq \varepsilon \} \leq 2 \exp \{ -k \min_{|t| \leq \varepsilon} G(\mu + t) \},
\] (29)

where \( G(a) = \sup_{\lambda}(a\lambda - \log \mathbb{E}e^{\lambda \xi}) = H(a) - a\theta + \psi(\theta) \). Let \( \phi(t) = G(\mu + t) \).

Then \( \phi(0) = 0, \phi'(0) = 0 \) and \( \phi''(t) = h'(\mu + t) = 1/\psi''(h(\mu + t)) > 0 \). Hence for some \( -\varepsilon \leq t_1 \leq \varepsilon \)

\[
\phi(\pm \varepsilon) = \varepsilon^2 \phi''(t_1)/2 \geq \varepsilon^2 \min_{|t| \leq \varepsilon} \phi''(t)/2 = \varepsilon^2 m/2.
\]

To apply this Lemma to our situation let \( \Delta \) be any closed subinterval of \( \Theta \), which contains \( \theta_1 \) and \( \theta_2 \) in its interior, and let \( m = \min \{1/\psi''(t), t \in \Delta \} = \min \{h'(x), x \in \psi'(\Delta) \} \). Also put \( M = \max \{1/\psi''(t), t \in \Delta \} = \max \{h'(x), x \in \psi'(\Delta) \} < \infty \).

We proceed with the proof of Proposition 3.1. Consider only the case \( \Lambda_\nu - \Lambda_k \) is at least \( n^\alpha \) as \( n \to \infty \) with the probability tending to 1.

Using (27) one obtains

\[
\Lambda_\nu - \Lambda_k = \nu H(\bar{x}_\nu) - kH(\bar{x}_k) + (n - \nu)H(\bar{x}_{\nu n}) - (n - k)H(\bar{x}_{kn})
\]

\[
= (\nu - k)H(\mu_1) + (n - \nu)H(\mu_2) - (n - k)H(\tilde{\mu}) + r_k
\]

\[
= \{tH(\mu_1) + (1 - t)H(\mu_2) - \tilde{H}(t\mu_1 + (1 - t)\mu_2) \}(n - k) + r_k;
\] (30)

where \( t = \frac{\nu - k}{\nu - \alpha} \) and \( \tilde{\mu} = \mathbb{E}\bar{x}_{kn} = t\mu_1 + (1 - t)\mu_2 \).

Let \( g(u) = uH(\mu_1) + (1 - u)H(\mu_2) - H(u\mu_1 + (1 - u)\mu_2) \). Since \( H \) is convex, one has \( g(u) \geq K_1 \min \{u, 1 - u\} \) for some \( K_1 > 0 \) and all \( 0 \leq u \leq 1 \). Indeed, it suffices to show that \( g(u)/u \) and \( g(u)/(1 - u) \) have strictly positive limits as \( u \downarrow 0 \) and \( u \uparrow 1 \) respectively. One has

\[
\lim_{u \downarrow 0} g(u)/u = g'(0) = H(\mu_1) - H(\mu_2) - H'(\mu_2)(\mu_1 - \mu_2) = H''(\tilde{\mu})(\mu_1 - \mu_2)^2/2 > 0
\]
where \( \bar{\mu} \) is some number between \( \mu_1 \) and \( \mu_2 \). Similarly, \( \lim_{u \uparrow 1} g(u)/(1 - u) = H''(\bar{\mu})(\mu_2 - \mu_1)^2/2 > 0 \) for some \( \bar{\mu} \) between \( \mu_1 \) and \( \mu_2 \). Hence, for \( n \) such that \( n^\alpha < \ln \omega \),

\[
\Lambda_\nu - \Lambda_k \geq K_1 \min\{t, 1 - t\}(n - k) + r_k = K_1(\nu - k) + r_k.
\]  

(31)

Consider the remainder \( r_k \). Since \( H'(x) = h(x) \) one derives from (30)

\[
\begin{align*}
 r_k &= \nu.H(\bar{x}_\nu) - H(\mu_1) - k.H(\bar{x}_k) - H(\mu_1) \\
 &\quad + (n - \nu).H(\bar{x}_\nu) - H(\mu_2) - (n - k).H(\bar{x}_k) - H(\mu_2) \\
 &= \nu.h(\mu_1)(\bar{x}_\nu - \mu_1) - k.h(\mu_1)(\bar{x}_k - \mu_1) \\
 &\quad + (n - \nu).h(\mu_2)(\bar{x}_\nu - \mu_2) - (n - k).h(\bar{\mu})(\bar{x}_k - \bar{\mu}) + r_k^{(2)}.
\end{align*}
\]

(32)

The error term \( r_k^{(2)} \) is estimated below.

As was observed, \( h(\mu_1) = \theta_j \). Also \( \bar{x}_\nu = [(\nu - k)\bar{x}_k + k\bar{x}_k]/\nu \) and \( \bar{x}_k = [(\nu - k)\bar{x}_k + (n - \nu)\bar{x}_\nu]/(n - k) \). Therefore

\[
\begin{align*}
 r_k &= (\nu - k)\theta_1(\bar{x}_k - \mu_1) + (n - \nu)\theta_2(\bar{x}_\nu - \mu_2) \\
 &\quad - h(\bar{\mu})[(\nu - k)(\bar{x}_k - \mu_1) + (n - \nu)(\bar{x}_\nu - \mu_2)] + r_k^{(2)} \\
 &= (\nu - k)(\theta_1 - h(\bar{\mu}))(\bar{x}_k - \mu_1) + (n - \nu)(\theta_2 - h(\bar{\mu}))(\bar{x}_\nu - \mu_2) + r_k^{(2)}.
\end{align*}
\]

(33)

In fact the second term of (33) has also the order \( \nu - k \). Indeed, the Taylor expansion of \( h(\bar{\mu}) \) shows that

\[
|(n - \nu)(\theta_2 - h(\bar{\mu}))| = |(n - \nu)h'(\zeta_1)(\mu_2 - \bar{\mu})| = (\nu - k)\frac{n - \nu}{n - k}|\mu_1 - \mu_2|h'(\zeta_1)
\]

\[
\leq (\nu - k)|\mu_1 - \mu_2| \max_{\zeta_1 \in \Theta(\Delta)} h'(\zeta_1) = (\nu - k)M|\mu_1 - \mu_2|
\]

with \( \zeta_1 \) between \( \mu_1 \) and \( \mu_2 \). Since \( h(x) \) is an increasing function, \( h(\zeta_1) \in \Delta \).

Similarly \( |\theta_1 - h(\bar{\mu})| \leq M|\mu_1 - \mu_2| \) and (33) shows that

\[
|r_k| \leq (\nu - k)M|\mu_1 - \mu_2||\bar{x}_k - \mu_1| + |\bar{x}_\nu - \mu_2| + |r_k^{(2)}| = |r_k^{(1)}| + |r_k^{(2)}|.
\]

Using Lemma 3 with \( \varepsilon = K_1/(6M|\mu_1 - \mu_2|) = K_2 \) one obtains for sufficiently large \( n \)

\[
P\left\{ |r_k^{(1)}| \geq K_1(\nu - k)/3 \right\} \leq P\left\{ M|\mu_1 - \mu_2|(\bar{x}_k - \mu_1) + |\bar{x}_\nu - \mu_2| \geq K_1/3 \right\}
\]

\[
\leq P\left\{ \max(|\bar{x}_k - \mu_1|; |\bar{x}_\nu - \mu_2|) \geq \frac{K_1}{6M|\mu_1 - \mu_2|} \right\} \leq 4\exp\{-mK_2^2n^\alpha/2\},
\]

(34)
The same upper bound can be obtained for the other terms of (36). Thus

\[ K \left. \text{with probability greater than } 1 - \frac{3}{8} \right| \]

(37)

where \( \zeta_2, \zeta_3, \zeta_4 \) and \( \zeta_5 \) are points in the corresponding intervals. Let us estimate the probability that all of them belong to the interval \( \psi'(\Delta) = h^{-1}(\Delta) \).

Denote by \( \delta_j \) the distance from \( \mu_j \) to the boundary of \( \psi'(\Delta) \) for \( j = 1, 2 \) and let \( \delta = \min(\delta_1, \delta_2) \). Then

\[ P \{ \zeta_2 \notin \psi'(\Delta) \} \leq P \{ |\zeta_2 - \mu_1| \geq \delta \} \leq P \{ |\tilde{x}_v - \mu_1| \geq \delta \} \leq 2 \exp\{ -m\delta^2 \nu/2 \}. \] (35)

The probabilities involving \( \zeta_3, \zeta_4 \) and \( \zeta_5 \) can be estimated similarly. So, if \( \gamma = \min(\nu, k, n - \nu, n - k) = \min(k, n - \nu) \geq \ln \omega \), then

\[ |r_k^{(2)}| \leq \frac{M}{2} |\nu(\tilde{x}_v - \mu_1) - k(\bar{x}_k - \mu_1)^2 + (n - \nu)(\tilde{x}_v - \mu_2)^2 - (n - k)(\bar{x}_k - \tilde{\mu})^2| \] (36)

with probability greater than \( 1 - 8 \exp\{ -m\delta^2 \nu/2 \} \). Applying Lemma 3 with \( \epsilon = K_1(\nu - k)/6M = K_3(\nu - k) \), we get

\[ P \left\{ \nu(\tilde{x}_v - \mu_1)^2 \geq K_3(\nu - k) \right\} = P \left\{ |\tilde{x}_v - \mu_1| \geq \sqrt{K_3(\nu - k)/\nu} \right\} \leq 2 e^{-mK_3n^\alpha/2}. \]

The same upper bound can be obtained for the other terms of (36). Thus

\[ P \left\{ |r_k^{(2)}| \geq K_1(\nu - k)/3 \right\} \leq 8 \exp\{ -m\delta^2 \nu/2 \} + \exp\{ -mK_3n^\alpha/2 \}. \]

By combining this result with (34) one obtains for \( n \to \infty \)

\[ P \left\{ |r_k| \geq \frac{2}{3} K_1(\nu - k) \right\} \leq 12 \exp\{ -K_4n^\alpha \} + o(\exp\{ -K_4n^\alpha \}), \] (37)

where \( K_4 = m \min(K_2^2, K_3)/2 \). This estimate and (31) show that with probability (37)

\[ \Lambda_\nu - \Lambda_k \geq K_1(\nu - k)/3 \geq K_1n^\alpha/3. \] (38)

By the definition of \( \hat{\nu} \), the event (38) implies that \( \Lambda_\nu - \Lambda_k \geq K_1n^\alpha/3 \), which shows that \( k \notin T_c \) for any \( n > (3c/K_1)^{1/\alpha} \). Thus as \( n \to \infty \) for \( K < K_1 \)

\[ P \{ [\ln \omega, \nu - n^\alpha] \cap T_c = \emptyset \} \leq 12 \nu \exp\{ -K_4n^\alpha \} (1 + o(1)) = o(\exp\{ -K_4n^\alpha \}). \]

By relabeling the data from \( X_n \) to \( X_1 \) one obtains a similar bound for \( \nu + n^\alpha \leq k \leq Ln \),

\[ P \{ [\nu + n^\alpha, n - \ln \omega] \cap T_c = \emptyset \} = o \left( e^{-K' n^\alpha} \right) \] as \( n \to \infty \)

with some positive constant \( K' \). These two results prove the proposition.
4.4 Proof of Proposition 3.2

Assume that $\nu - n^\alpha \leq k \leq \nu$ and consider the function $q(\theta, x) = \log f(x|\theta) = \theta x - \psi(\theta)$ for which $H(x) = q(h(x), x)$. By the definition of $S_k$

$$S_\nu - S_k = \nu q(\theta_1, \bar{x}_\nu) - kq(\theta_1, \bar{x}_k) + (n - \nu)q(\theta_2, \bar{x}_m) - (n - k)q(\theta_2, \bar{x}_k),$$

and (30) implies that

$$(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k) = \nu[H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu)] - k[H(\bar{x}_k) - q(\theta_1, \bar{x}_k)] + (n - \nu)[H(\bar{x}_m) - q(\theta_2, \bar{x}_m)] - (n - k)[H(\bar{x}_k) - q(\theta_2, \bar{x}_k)] = A_1 + A_2 + A_3,$$

where

$$A_1 = (\nu - k) \{[H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu)] - [H(\bar{x}_m) - q(\theta_2, \bar{x}_m)]\};$$

$$A_2 = k \{H(\bar{x}_\nu) - H(\bar{x}_k) - [q(\theta_1, \bar{x}_\nu) - q(\theta_1, \bar{x}_k)]\};$$

$$A_3 = (n - k) \{H(\bar{x}_m) - H(\bar{x}_k) - [q(\theta_2, \bar{x}_m) - q(\theta_2, \bar{x}_k)]\}.$$

Consider these terms separately. The Taylor expansion of the function $q(\theta, x)$ in the first variable gives

$$H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu) = q(h(\bar{x}_\nu), \bar{x}_\nu) - q(\theta_1, \bar{x}_\nu) = \frac{\partial q}{\partial \theta}(\theta_1, \bar{x}_\nu)(h(\bar{x}_\nu) - \theta_1) + \frac{1}{2} \frac{\partial^2 q}{\partial \theta^2}(\zeta_6, \bar{x}_\nu)(h(\bar{x}_\nu) - \theta_1)^2$$

for some $\zeta_6$ between $\theta_1$ and $h(\bar{x}_\nu)$. Here $\frac{\partial q}{\partial \theta}(\theta_1, \bar{x}_\nu) = \bar{x}_\nu - \psi'(\theta_1) = \bar{x}_\nu - \mu_1$ and $\frac{\partial^2 q}{\partial \theta^2}(\zeta_6, \bar{x}_\nu) = -\psi''(\zeta_6)$. The Taylor formula for $h(x)$ shows that

$$H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu) = \theta h'(\zeta_7)(\bar{x}_\nu - \mu_1)^2 - \psi''(\zeta_6) h'(\zeta_7)(\bar{x}_\nu - \mu_1)^2 / 2 = \theta h'(\zeta_7)(1 - \psi'(\zeta_6))^2 (\bar{x}_\nu - \mu_1)^2,$$

and a similar expression holds for $H(\bar{x}_m) - q(\theta_1, \bar{x}_m)$.

As in the proof of Proposition 3.1, $\zeta_6 \in \Delta$ and $\zeta_7 \in \psi'(\Delta)$ with the probability $o(\exp\{-Cn^\alpha\})$ for some $C > 0$. Then with the probability of the same order

$$|A_1| \leq (\nu - k)M (1 + M/2m) \left\{(\bar{x}_\nu - \mu_1)^2 + (\bar{x}_m - \mu_2)^2\right\} \leq 2n^\alpha M (1 + M/2m) \max\left[(\bar{x}_\nu - \mu_1)^2; (\bar{x}_m - \mu_2)^2\right].$$

19
Hence by Lemma 3

\[
P\left\{|A_1| \geq \frac{\varepsilon_n}{3}\right\} \leq P\left\{\max(|\vec{x}_\nu - \mu_1|, |\vec{x}_\nu n - \mu_2|) \geq \sqrt{\frac{\varepsilon_n/3}{2n^\alpha M (1 + M/2m)}}\right\}
= o\left(\exp\left\{-Cn^{\omega-\alpha-\beta}\right\}\right)
\]

as \( n \to \infty \) for some \( C > 0 \).

The second term of (39) can be estimated as follows

\[
A_2 = k \left\{H(\vec{x}_\nu) - H(\vec{x}_k) - \theta_1(\vec{x}_\nu - \vec{x}_k)\right\}
\leq k \left\{h(\vec{x}_k)(\vec{x}_\nu - \vec{x}_k) + h'(\zeta_8)(\vec{x}_\nu - \vec{x}_k)^2/2 - \theta_1(\vec{x}_\nu - \vec{x}_k)\right\}
\leq k(\vec{x}_\nu - \vec{x}_k) \left\{h'(\zeta_8)(\vec{x}_\nu - \vec{x}_k)/2 + h'(\zeta_9)(\vec{x}_k - \mu_1)\right\}.
\]

Here, similarly to (35), with probability \( 1 - o\left(e^{-Cn^\omega}\right) \) the intermediate values \( \zeta_8 \) and \( \zeta_9 \) belong to \( \psi'(\Delta) \) and

\[
|A_2| \leq 1.5 n^\alpha |\vec{x}_{k\nu} - \vec{x}_k| M \max(|\vec{x}_\nu - \vec{x}_k|, |\vec{x}_k - \mu_1|).
\]

Let \( 0 < \sigma < \omega - \alpha - \beta \). By Lemma 3

\[
P\left\{\max(|\vec{x}_\nu - \vec{x}_k|, |\vec{x}_k - \mu_1|) \geq \frac{\varepsilon_n}{4.5 Mn^{\alpha+\sigma}}\right\} = o\left(\exp\left\{-Cn^{\omega-2(\alpha+\beta+\sigma)}\right\}\right)
\]

as \( n \to \infty \) for some positive \( C \). Also

\[
P\left\{|\vec{x}_{k\nu} - \vec{x}_k| \geq n^\sigma\right\} \leq P\{|\vec{x}_{k\nu}| \geq n^\sigma/2\} + P\{|\vec{x}_k| \geq n^\sigma/2\} \leq 2P_{\theta_1}\{|X_1| \geq n^\sigma/2\},
\]

which is \( o\left(\exp\{-Cn^\sigma\}\right) \) for some \( C \). Indeed, for sufficiently small \( \lambda \)

\[
E_{\theta_1} e^{\lambda |X_1|} \leq E_{\theta_1} e^{\lambda X_1} + E_{\theta_1} e^{-\lambda X_1} = e^{-\psi(\theta_1)} \left(e^{\psi(\theta_1+\lambda)} + e^{\psi(\theta_1-\lambda)}\right) < \infty
\]

and

\[
P_{\theta_1}\{|X_1| \geq n^\sigma/2\} = \int_{|x| \geq n^\sigma/2} f(x|\theta_1)d\mu(x)
\leq \exp\{-\lambda n^\sigma/2\} \int_{|x| \geq n^\sigma/2} e^{\lambda|x|} f(x|\theta_1)d\mu(x) \leq E_{\theta_1} e^{\lambda |X_1|} \exp\{-\lambda n^\sigma/2\}.
\]

Hence, for any \( 0 < \sigma < \omega/2 - \alpha - \beta \),

\[
P\left\{|A_2| \geq \varepsilon_n/3\right\} = o\left(\exp\left\{-Cn^{\omega-2(\alpha+\beta+\sigma)}\right\} + \exp\{-Cn^\sigma\}\right).
\]
We choose \( \sigma = (\omega - 2\alpha - 2\beta)/3 \), in which case
\[
P\left\{ |A_2| \geq \varepsilon_n/3 \right\} = o\left( \exp\left\{ -Cn^{(\omega-2\alpha-2\beta)/3} \right\} \right).
\]
The same estimate holds for \( P\{ |A_3| \geq \varepsilon_n/3 \} \). Finally, by (39) and (40)
\[
P\left\{ |(\Lambda_{\nu} - \Lambda_k) - (S_{\nu} - S_k)| \geq \varepsilon_n \right\} = o\left( \exp\left\{ -Cn^{(\omega-2\alpha-2\beta)/3} \right\} \right).
\]
The case \( \nu < k < \nu + n^\alpha \) can be treated similarly, and one has
\[
P\left\{ \max_{|\nu-n^\alpha,\nu+n^\alpha|} \left| (\Lambda_{\nu} - \Lambda_k) - (S_{\nu} - S_k) \right| \geq \varepsilon_n \right\} = o\left( \exp\left\{ -Cn^{(\omega-2\alpha-2\beta)/3} \right\} \right).
\]

4.5 Proof of Theorem 3.3

Let \( \alpha \) be an arbitrary number in \((0, \omega/2 - \beta)\). Then for any \( k \leq \nu - n^\alpha \)
\[
S_{\nu} - S_k = \sum_{j=k+1}^{\nu} \log \frac{f(x_j|\theta_1)}{f(x_j|\theta_2)} = (\nu - k)\{ \rho_1 + (\theta_1 - \theta_2)(\bar{x}_{\nu} - \mu_1) \}.
\]

It follows from Lemma 3 that
\[
1 - P\left\{ |\bar{x}_{\nu} - \mu_1| \leq \frac{\rho_1}{2|\theta_1 - \theta_2|} \right\} = o\left( e^{-Cn^\alpha} \right)
\]
for some positive \( C \). Hence, \( S_{\nu} - S_k \geq S_{\nu} - S_k \geq \rho_1 n^\alpha/2 > c \) with probability
\[
1 - o(e^{-Cn^\alpha})
\]
for any \( n > (2c/\rho_1)^{1/\alpha} \), which implies \( k \not\in R_c \). The case \( k \geq \nu + n^\alpha \) is similar. Thus
\[
P\left\{ R_c \not\subseteq [\nu - n^\alpha, \nu + n^\alpha] \right\} = P\{ \min (S_{\nu} - S_k , k \not\in [\nu - n^\alpha, \nu + n^\alpha]) \leq c \} = o\left( e^{-Cn^\alpha} \right). \tag{42}
\]

Here \( c \) is any bounded function of \( n \), so that it can be replaced by \( c - \varepsilon_n \). By
\[
(42) \text{ and Proposition 3.1, } R_{c - \varepsilon_n} \text{ and } T_c \text{ are both subsets of } [\nu - n^\alpha, \nu + n^\alpha] \text{ with probability } 1 - o(e^{-Cn^\alpha}) \text{ for some } C > 0. \text{ Also by Proposition 3.2}
\]
\[
P\left\{ \max_{\nu - n^\alpha \leq k \leq \nu + n^\alpha} \left| (\Lambda_{\nu} - \Lambda_k) - (S_{\nu} - S_k) \right| \geq \varepsilon/2 \right\} = o\left( \exp\left\{ -Cn^{(\omega-2\alpha-2\beta)/3} \right\} \right).
\]
For any $k$ in $R_{c-\varepsilon_n}$, $S_{\hat{\nu}} - S_k \leq c - \varepsilon_n$ and

$$
\Lambda_{\hat{\nu}} - \Lambda_k = \{(\Lambda_{\nu} - \Lambda_k) - (S_{\nu} - S_k)\} + \{(\Lambda_{\hat{\nu}} - \Lambda_\nu) - (S_{\hat{\nu}} - S_\nu)\} + (S_{\hat{\nu}} - S_k)
\leq \varepsilon_n/2 + \varepsilon_n/2 + (c - \varepsilon_n) = c
$$

(43)

with probability $1 - o\left(\exp\left\{\frac{-Cn(\omega - 2\beta)}{3}\right\}\right)$. But $\Lambda_{\hat{\nu}} - \Lambda_k \leq c$ means that $k \in T_c$, so that $R_{c-\varepsilon_n} \subset T_c$ with the probability of the same order for any $0 < \alpha < \omega/2 - \beta$. The optimal choice is, $\alpha = (\omega - 2\beta)/5$, in which case

$$
P\{R_{c-\varepsilon_n} \subset T_c\} = 1 - o\left(\exp\left\{-Cn(\omega - 2\beta)/5\right\}\right).
$$

The other inclusion, $T_c \subset R_{c+\varepsilon_n}$, can be proven similarly by using (43).

References


