APPROVAL SHEET

Title of Dissertation: Confidence Estimation in the Change-Point Problem

Name of Candidate: Michael I. Baron

Doctor of Philosophy, 1995

Dissertation and Abstract Approved: ______________________

Dr. Andrew L. Rukhin

Professor

Department of Mathematics and Statistics

Date Approved: ______________________

CURRICULUM VITAE

Name: Michael I. Baron.

Permanent address: Department of Mathematics & Statistics, UMBC.

Degree and date to be conferred: Ph.D., 1995.

Date of birth: March 31. 1968.

Place of birth: Leningrad, Russia.

Secondary education: School No 239 of St. Petersburg.

Collegiate institutions attended	Dates	Degree	Date of degree
University of Maryland	1992-1995	Ph.D.	1995
St. Petersburg State University, Russia	1985–1986, 1988–1992	M.Sc.	1992

Major: Statistics.

Professional publications:

1. M.Baron, A.L.Rukhin. Asymptotic behavior of confidence regions in the change-point problem (*submitted*).
2. M.Baron. On the first passage time for queuing processes (*submitted*).

Professional positions held:

1995-1995	Instructor, Department of Mathematics & Statistics, UMBC.
1992-1995	Graduate Assistant, Department of Mathematics & Statistics, UMBC.

ABSTRACT

Title of Dissertation: Confidence Estimation in the Change-Point Problem

Michael I. Baron, Doctor of Philosophy, 1995

Dissertation directed by: Dr. Andrew L. Rukhin, Professor

Department of Mathematics and Statistics

In the classical setting of the change-point problem, the traditional confidence regions for estimating the change-point parameter are considered. For the case, when the distributions before and after the change are completely known, the asymptotic behavior of the confidence sets is derived in terms of the coverage probability and the expected width. These results allow to obtain the optimal form of the confidence regions for the large sample sizes. As a measure of quality, we assume an appropriate loss function, which takes into account the number of elements and efficiency of confidence estimators. An example of Bernoulli distributions is considered, and the exact formulae for this case are derived.

If pre-change and after-change distributions are unknown, it is assumed that they belong to a one-parameter exponential family. The one- or multidimensional nuisance parameter, which changes at an unknown moment, is estimated by the maximum likelihood procedure for every hypothetical value of the change point. With these estimators, the confidence regions for the change-point parameter are defined by analogy with those which were used for known distributions. Similar results concerning the coverage probability, the expected width, and therefore, the risk function of these sets are obtained.

The study of convergence rates of the main characteristics of the confidence sets leads to interesting problems of renewal theory, related to the renewal measure and the first passage time. These problems are discussed in the general case and in the case of Bernoulli distributions, where the exact asymptotics is derived. The probability of the correct decision, the coverage probability and the expected width are shown to converge exponentially fast as the sample size increases to infinity. Thus, the limiting expressions obtained in the first chapters can be also used for detecting a change point in finite, although sufficiently large, samples.

As an applied example, the data of measurements of mass standards, released by the National Institute of Standards and Technology (NIST), is studied.

CONFIDENCE ESTIMATION IN THE CHANGE-POINT PROBLEM

by
Michael I. Baron

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1995

This thesis is dedicated

to my parents

ACKNOWLEDGMENT

The author will be always indebted to his research advisor Dr. Andrew L. Rukhin for invaluable help and unlimited enthusiasm in advising him. He is also deeply thankful to Dr. Rukhin for the variety of graduate courses that he was teaching. Next, he would like to thank the other professors of the Department of Mathematics and Statistics at the University of Maryland, Baltimore County, and especially Dr. Bimal K. Sinha, Dr. Thomas Mathew, Dr. Peter C. Matthews and Dr. Neerchal K. Nagaraj, for their extended help and interesting courses, from which he has benefited a lot. He also thanks Dr. Matthews for his comments and significant corrections of this thesis. Very special thanks are due to the department chairman Dr. Rouben Rostamian for creating such a nice academic atmosphere. Also, in this regard, the author is grateful to Mrs. Sara Anderson for her help during his three years at UMBC. Finally, he would never forget to thank his friends, especially his fiancee Ms. Lyudmila Kravets, Mr. Andrew Yershov and Mr. Ilya Dondoshansky, for their permanent support, help and encouragement.

Contents

List of Figures

Chapter 1

Introduction

1.1 Estimating the "disorder" time

There are many practical situations when the system goes out of control or signals a change, and it is necessary to detect and to estimate the "disorder" time. Meteorological data appeared to be significant to detect an evident change in the pattern of the Nile river overflows near the city of Aswan, Egypt, in the end of the last century (see Cobb (1978)). MacNeill and Mao (1993) studied a rapid rise in the number of reported Kaposi's Sarcoma diagnosed cases. The same paper examines a change in female mortality rates for pre-menopausal breast cancer. Telephone companies usually cancel calling cards if they notice a significant change in a Poisson process of telephone calls. Berry and Hartigan (1993) conclude that only allowing for changing team strengths with time during the baseball season is it possible to describe the situation in the National League and to make trustworthy predictions. In chapter 5 of this thesis we consider a change in the check standard data from the mass calibration process.

Identification of changes at unknown time and estimation of the location of changes are referred to as change-point problem. Although it is believed that this problem goes back to A. N. Kolmogorov, one of the oldest references is Shewhart (1931). For a long period in the past Shewhart's $\bar{X}$-Charts were very popular in detecting shifts in the mean of a production process.

However, this procedure has been found to be inefficient in detecting small changes. Page (1954, 1955, 1957) proposed more sensible methods for identifying small shifts in the mean of the process and testing the hypothesis of "no change". They were further developed in Khan (1978, 1979a), and they are known as the cumulative sum (CUSUM) procedures. Based on the observations $X_1, X_2, \ldots$, a random walk $S_k = \sum_{j=1}^{k} X_j$ is defined, and the corresponding queueing process

$$W_k = S_k - \min_{0 \le j \le k} S_j$$

is considered. The CUSUM procedure rejects a hypothesis of "no change" on the given interval, if W_k attains large values on it. The first passage time $\inf\{k,\ W_k > h\}$ for some fixed h is an optimal stopping rule, i.e. the procedure signals a change in the mean when W_k exceeds level h. At that moment the process is said to be out of control. In a more general framework, in order to identify changes in distribution, other than mean shifts, the random walk S_k is based on the likelihood ratio of pre-change and after-change distributions, which is clearly a sufficient statistic for the change-point parameter. The likelihood ratio test and its modifications were also used for testing the "no change" hypothesis.

The point estimation of the change point was initiated by Chernoff and Zacks (1964) (Bayesian approach) and Hinkley (1970) (maximum likelihood procedure). Most of the papers on this subject considered estimating the location of a mean shift in a sequence of independent normal random variables. According to the "Bayesian-type" approach of H. Chernoff and S. Zacks an appropriate prior distribution for the nuisance parameters associated with both the null and the alternative hypotheses was assumed, then the unconditional likelihood functions were obtained through elimination of the nuisance parameters, and finally the likelihood ratio statistics was derived. D. Hinkley was the first to point out the key role of the extremum of a two-sided random walk in the context of maximum likelihood estimation of the change point. In this thesis, we widely use the ideas of both papers for confidence estimation of the change-point parameter.

Much was done in the field of parametric change-point estimation. Smith (1975) and Worsley (1986) studied the problem in the exponential family framework. Only during the last two decades the problem was studied in its nonparametric setting, that is, when there is no prior information about the distributions before and after the change. The two standard references here are Darkhovskhi (1976) and Carlstein (1988). Brodsky and Darkhovsky (1993) summarize nonparametric procedures for estimating the change point. Most of these methods imply a comparison of empirical distribution functions before and after the change for every hypothetical value of the change-point parameter under various measures of discrepancy between the distributions. The time moment which maximizes the discrepancy is used as a nonparametric change-point estimator.

1.2 Confidence regions for the change-point parameter

Following the classical setting of the change-point problem, we assume that the vector of observed data $\mathbf{X} = (X_1, \ldots, X_\nu, X_{\nu+1}, \ldots, X_n) = (\mathbf{X}_1, \mathbf{X}_2)$ consists of two independent parts, the first $\mathbf{X}_1$ being a random sample from distribution F, and the second random subsample $\mathbf{X}_2$ coming from distribution G. The change point ν, whose value is unknown, is a parameter of interest. The distributions F and G are assumed to have a common support. Let f and g be the corresponding densities with respect to some dominating probability measure.

We study the confidence estimation of the change-point parameter in a decision theoretic aspect. An appropriate risk function for a set C would combine its expected width and coverage probability. Thus, we assume a loss function

$$W(\nu, C) = \lambda|C| + 1 - I_C(\nu), \tag{1.1}$$

where $|C|$ is the cardinality of a confidence set, I denotes the indicator function and λ is a fixed positive constant. It is known, that there is no consistent estimator in our setting, and it becomes more difficult to detect a change point in large samples. Therefore, for large samples one needs relatively large confidence regions for the change-point parameter, and this can be reflected in the proposed loss function if we let $\lambda \to 0$ as $n \to \infty$.

Five different confidence sets for estimating the change-point parameter are discussed in Siegmund (1988). For completely known distributions F and G, the classical credible confidence region for ν has the form

$$R_c = \left\{ k : \max_i \left[\sum_1^i \log f(X_j) + \sum_{i+1}^n \log g(X_j) \right] < \sum_1^k \log f(X_j) + \sum_{k+1}^n \log g(X_j) + c \right\}$$

for some positive c. Another, less popular set is an interval modification of R_c, and it is defined as $[\inf R_c;\ \sup R_c] = [l;\ r]$. For a particular case of normal distributions with mean which changes from 0 to $\delta > 0$ at the time ν and common unit variance, the asymptotics of the expected width for these two sets is evaluated in Siegmund (1988) as follows,

$$\mathbf{E}|R_c| = \frac{4c}{\delta^2} + \frac{4}{\delta^2} - \frac{4}{\delta} \int_0^\infty \{2\ \mathbf{P}_0(M > x) - \ \mathbf{P}_0^2(M > x)\}\, dx + o(1),$$

and

$$\mathbf{E}(r - l) = \frac{4c}{\delta^2} + \frac{4}{\delta^2}$$
$$-\frac{4}{\delta} \int_0^\infty \int_0^\infty \ \mathbf{P}_0(M \in dy)\{2\ \mathbf{P}_0(M > x + y) - \ \mathbf{P}_0^2(M > x + y)\}\, dx + o(1),$$

as $c \to \infty$, where $M = \sup_{k\geq 0} \sum_{j=1}^{k} \log f(X_j)/g(X_j)$.

One can also consider another interval, the connected component of the set R_c which contains the maximum likelihood estimator of the change point

$$\hat{\nu} \;=\; \arg\max_{1\leq k\leq n} \sum_{j=1}^{k} \log \frac{f(X_j)}{g(X_j)}.$$

It will follow from further results that the conditional probability of covering the true value of ν by this interval given $\nu \in R_c$ converges to 1 exponentially fast, so that including the other components of R_c will most probably increase the risk.

The Bayesian procedure results in quite a different confidence set. Assuming the uniform prior distribution of the parameter ν on the set $\{1,\ 2,\ \ldots,\ n\}$, it is desired to minimize

$$B(C) = \mathbf{E}\,\mathbf{E}_\nu\; W(\nu, C) = \mathbf{E}\,(\lambda\;\mathbf{E}_\nu|C| + 1 - \mathbf{P}_\nu\{\nu \in R_c\})$$

or

$$\begin{aligned} n\{B(C)-1\} &= \sum_{\nu=1}^{n} \int \sum_{k\in C} (\lambda - I_\nu(k)) f(x_1)\ldots f(x_\nu) g(x_{\nu+1})\ldots g(x_n)\, dx_1 \ldots dx_n \\ &= \mathbf{E}_G \sum_{k\in C} \left(\lambda \sum_{\nu=1}^{n} \xi_1 \ldots \xi_\nu - \xi_1 \ldots \xi_k \right), \end{aligned}$$

where $\xi_j = f(X_j)/g(X_j)$ for $j = 1\ldots n$. Hence the Bayes rule has the form

$$C_B = \left\{ k:\ \xi_1 \ldots \xi_k \;>\; \lambda \sum_{\nu=1}^{n} \xi_1 \ldots \xi_\nu \right\}.$$

On the other hand, the set R_c introduced above can be represented as

$$R_c = \left\{ k:\ \xi_1 \ldots \xi_k \;>\; e^{-c} \max_{1\leq\nu\leq n} \xi_1 \ldots \xi_\nu \right\}.$$

It follows from (2.23) that the value of c that minimizes the risk (and also the Bayes risk with respect to the uniform prior) under the loss function (1.1) is asymptotically proportional to $-\log\lambda$ as $\lambda \to 0$ and $c \to \infty$. Then, with the optimal choice of

c, one has $e^{-c} \sim \lambda$, and therefore, the set R_c is defined similarly to the Bayes procedure if one replaces $\sum_{\nu=1}^{n} \xi_1 \ldots \xi_\nu$ by $\max_{1 \leq \nu \leq n} \xi_1 \ldots \xi_\nu$.

We derive the asymptotic expressions for the coverage probability and the expected width of the confidence set R_c in Chapter 2. The limiting behavior of the risk, which does not depend on the parameter ν and therefore coincides with the Bayes risk, follows from these results.

The situation of partially known pre-change and after-change distributions is discussed in Chapter 3. That is, we assume that they belong to an exponential family whose parameter changes at the time ν, but remains unknown. For this case we define the confidence region for estimating the change-point parameter by analogy with the set R_c and derive its asymptotic properties.

Chapter 4 studies the rates of convergence of the coverage probability and the expected width of the defined sets as the sample size increases to infinity. A practical illustration is given in Chapter 5 where we study the point and confidence estimation of the change point in the calibration process of measurements of mass standards.

Chapter 2

The case of known distributions

To study the confidence estimation problem described in Chapter 1 we consider the following asymptotic setting originally suggested by Cobb (1978). Let $\{X_j,\ j \in \mathbf{Z}\}$ be a doubly infinite sequence of independent random variables. Assume that for some integer ν, 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 setting to the situation described in Chapter 1, with $m = [(n-1)/2]$ 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 $-m \leq \nu \leq [n/2]$.

Let

$$Z_j = \log \frac{f(X_j)}{g(X_j)} \ \text{ for } \ -\infty < j < +\infty;$$

$$S_k = \sum_{j=1}^{k} Z_j \ \text{ for } k \geq 1;$$

$$S_k = -\sum_{j=k+1}^{0} Z_j \ \text{ for } \ k \leq -1;$$

$$S_0 = 0.$$

Then the credible confidence set introduced in Chapter 1 has the form

$$R_c = \{k : \ S_k < \max_j S_j - c\} \tag{2.1}$$

for some positive constant c.

Since

$$\begin{aligned} \mathbf{E}\ Z_j &= K(G,F) = \rho_1 > 0 \text{ for } j \le \nu \text{ and} \\ \mathbf{E}\ Z_j &= -K(F,G) = -\rho_2 < 0 \text{ for } j > \nu \end{aligned} \tag{2.2}$$

with K denoting Kullback-Leibler information numbers, according to the strong law of large numbers $\max_j S_j$ is a proper random variable, so that R_c is defined correctly. Throughout this paper, if Z_1 is an arithmetic random variable with a span d, that is, if $Z_1 \in \{..., -2d, -d, 0, d, 2d, ...\}$ with probability 1, and d is the largest constant with this property, we let c tend to $+\infty$ only through the multiples of d.

Our immediate goal is to derive the asymptotic formulae for the coverage probability and the expected width of R_c. In our setting of the problem $\mathbf{P}_\nu\{\nu \in R_c\}$ and $\mathbf{E}_\nu|R_c|$ do not depend on ν, so we can look at the case $\nu = 0$ only. Then one can consider $\{S_k,\ k \in \mathbf{Z}\}$ as a combination of two independent random walks:

$$S_k^{(1)} = S_{-k} \text{ and } S_k^{(2)} = S_k \text{ for } k \ge 0.$$

In what follows we attach the index 1 or 2 to all the probabilities and expected values related to $S_k^{(1)}$ and $S_k^{(2)}$ respectively.

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

$$\tau^{(j)}(x) = \begin{cases} \inf\{k : S_k^{(j)} \ge x\} \text{ for } x \ge 0; \\ \inf\{k : S_k^{(j)} \le x\} \text{ for } x < 0; \end{cases}$$

$$\tau_+^{(j)} = \inf\{k : S_k^{(j)} > 0\}; \quad \tau_-^{(j)} = \inf\{k : S_k^{(j)} < 0\}$$

for $j = 1, 2$.

Theorem 2.1 *For $c \to \infty$*

$$1 - \mathbf{P}_\nu\{\nu \in R_c\} \sim e^{-c}\eta_1\eta_2'\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right),$$

where

$$\eta_j = \exp\left\{-\sum_{k=1}^{\infty} k^{-1}\,\mathbf{P}_j(S_k > 0)\right\}, \quad \eta_j' = \exp\left\{\sum_{k=-\infty}^{-1} k^{-1}\,\mathbf{P}_j(S_k \geq 0)\right\} \tag{2.3}$$

for $j = 1, 2$, and ρ_1, ρ_2 are information numbers.

Proof:

We can think of f and g as embedded into one-parameter exponential family of the form

$$f(x|\theta) = \exp\left\{\theta \log \frac{f}{g}(x) - \psi(\theta)\right\} f(x|0).$$

Here the reference measure is G and $\exp\{\psi(\theta)\} = \int f^\theta g^{1-\theta}$ so that $f(x|1) = f(x)$ and $f(x|0) = g(x)$. Moreover, one has $\psi'(0) = -\rho_2$ and $\psi'(1) = \rho_1$. Since

$$\mathbf{P}\{\nu \in R_c\} = \mathbf{P}\{M_1 < c\}\;\mathbf{P}\{M_2 < c\} = \mathbf{P}_1\{\tau(c) = \infty\}\;\mathbf{P}_2\{\tau(c) = \infty\}, \tag{2.4}$$

we can make use of formula (8.48) of Siegmund (1985) which evaluates

$$\mathbf{P}_1\{\tau(c) < \infty\} = 1 - \mathbf{P}_1\{\tau(c) = \infty\}$$

showing that for $c \to \infty$

$$\mathbf{P}_1\{\tau(c) < \infty\} \sim e^{-c}\,\mathbf{P}_1\{\tau_+ = \infty\}\;\mathbf{P}_2\{\tau_- = \infty\}/\psi'(1) = e^{-c}\eta_1\eta_2'/\rho_1. \tag{2.5}$$

Corollary 8.44 of Siegmund (1985) was used here to evaluate $\mathbf{P}_1\{\tau_+ = \infty\}$ and $\mathbf{P}_2\{\tau_- = \infty\}$.

Similar asymptotics obtains for $\mathbf{P}_2\{\tau(c) < \infty\}$ so that Theorem 2.1 follows from (2.4).

□

A similar result for point estimation of the change-point parameter for the zero-one loss function is given in Rukhin (1994).

An immediate use of Theorem 2.1 is that it allows to construct an approximately $(1-\alpha)100\%$ confidence set. One only has to take R_c with

$$c \geq -\log\alpha + \log\eta_1\eta_2'\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right). \tag{2.6}$$

As an example we consider the situation when F and G are Bernoulli distributions with probabilities of success $p,\ 0 < p < 1/2$, and $(1-p)$ respectively. In that case it is possible to obtain the exact expression for the coverage probability.

According to (2.4), by a translation invariant property of $\mathbf{P}_\nu\{\nu \in R_c\}$,

$$\mathbf{P}_\nu\{\nu \in R_c\} = \mathbf{P}_0\{0 \in R_c\} = \mathbf{P}_F\{\max_{k\leq 0} S_k < c\}\,\mathbf{P}_G\{\max_{k\geq 0} S_k < c\} = \tilde{Q}_1(c)\tilde{Q}_2(c).$$

Here $\tilde{Q}_j(c)$ is the left-continuous version of the distribution function of M_j for $j = 1, 2$. It satisfies to the following integral equation

$$\tilde{Q}_j(c) = \int \mathbf{P}_j\{S_1 \in du\}\tilde{Q}_j(c-u) = \mathbf{E}_j\tilde{Q}_j(c-S_1) \quad \text{for} \quad c \geq 0. \tag{2.7}$$

As to the boundary conditions, $\tilde{Q}_j(c)$ vanishes for non-positive c and monotonically increases to 1 as c increases to infinity.

The exact solution to (2.7) can be found for the case of Bernoulli distributions. Indeed, in that case $\tilde{Q}_1(c) = \tilde{Q}_2(c) = \tilde{Q}(c)$, and S_1 takes values ζ and $-\zeta$ with probabilities p and $1-p$, where $\zeta = \log\frac{1-p}{p}$. Hence $\max_{k\geq 0} S_k$ can take only the values which are multiples of ζ, and the function $\tilde{Q}(c)$ is a constant on any interval

$(n\zeta, (n+1)\zeta]$ for integer n. Put $c = n\zeta$ and $a_n = \tilde{Q}(c)$. Then (2.7) implies the following difference equation for a_n:

$$\begin{array}{ll} a_n = pa_{n-1} + (1-p)a_{n+1}, & n \in \mathbf{Z}_0 \\ a_n = 0, & n \leq 0 \\ a_n \uparrow 1, & n \uparrow +\infty. \end{array}$$

Its unique solution can be found, say, by the method of moment generating functions. It has the form $a_n = 1 - \exp\{-n\zeta\}$, $n > 0$, so that

$$\tilde{Q}(c) = a_{c/\zeta} = 1 - e^{-c}, \tag{2.8}$$

and

$$\mathbf{P}_\nu\{\nu \in R_c\} = \left(1 - e^{-c}\right)^2.$$

According to this formula, for $c \geq -\log(1 - \sqrt{1-\alpha})$, R_c is a $(1-\alpha)100\%$ parameter-free confidence set for the change point ν.

We also note that setting $c = \zeta$ in (2.8) one can obtain an expression for $\eta = \exp\{-\sum_{k=1}^{\infty} k^{-1}\, \mathbf{P}(S_k > 0)\}$ for the case of Bernoulli distributions,

$$\eta \;=\; \mathbf{P}\{\tau_+ = \infty\} \;=\; \mathbf{P}\{\max_{k\geq 0} S_k < \zeta\} \;=\; 1 - e^{-\zeta} \;=\; 1 - \frac{p}{q}. \tag{2.9}$$

The next result pertains to the asymptotics of the expected width.

Theorem 2.2 *Let the distributions F and G satisfy to one of the following conditions:*

(i) F and G are continuous distributions.

(ii) F and G are mixed distributions (i.e. they have both discrete and continuous components), and

$$\textstyle\sum^* e^x\, \mathbf{P}\{S_{\tau+} = x\} < 1, \tag{2.10}$$

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

(iii) F and G are discrete distributions, $\mathbf{E}_F(\frac{f}{g})^\epsilon$ and $\mathbf{E}_G(\frac{g}{f})^\epsilon$ are both finite for some $\epsilon > 0$, and (2.10) holds.

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

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

where b is a constant term specified in (2.22).

Proof:

In the notations above,

$$\begin{aligned} \mathbf{E}|R_c| &= \int_{u\geq 0} \mathbf{E}\left\{|R_c| \mid M = u\right\} dQ(u) \\ &= \int_{u\geq 0} \mathbf{E}\left\{|R_c| \mid M_1 = u;\ M_2 \leq u\right\} Q_2(u)\, dQ_1(u) \\ &+ \int_{u\geq 0} \mathbf{E}\left\{|R_c| \mid M_1 < u;\ M_2 = u\right\} Q_1(u^-)\, dQ_2(u), \end{aligned} \tag{2.11}$$

where $Q_1(u^-)$ denotes the left-hand limit of $Q_1(u_1)$ as $u_1 \to u$. We consider only the first integral in the right-hand side of (2.11) because the second integral can be evaluated similarly.

Since R_c may be disconnected (i.e. it can have lacunas) we 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 $H_0,\ H_1,\ \ldots$. In this notation, we put $G_0 = \emptyset$ if $0 \notin R_c$, so that G_0 contains 0 if only it is not empty. Conversely, H_0 does not contain 0, that is, $H_0 \subset (0, \infty)$, and similarly to the definition of G_0, $H_0 = \emptyset$ 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.

Hence, for any positive u

$$\mathbf{E}\left\{|R_c| \mid M_1 = u; M_2 \le u\right\} \tag{2.12}$$
$$= \mathbf{E}\left\{|G_0| + \sum_{i=1}^{n_1} |G_i| \mid M_1 = u \ge M_2\right\} + \mathbf{E}\left\{|H_0| + \sum_{i=1}^{n_2} |H_i| \mid M_2 \le u = M_1\right\}.$$

Considering separately the four terms that appear in (2.12), let

$$G_{u,c} = \left\{k \le \tau(u-c) \ : \ S_k^{(1)} > u - c\right\}, \ \text{if} \ u \le c\ ;$$
$$G_{u,c} = \emptyset, \ \text{otherwise.}$$

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

$$\begin{aligned}\mathbf{E}\{|G_0| \mid M_1 = u \ge M_2\} &= \mathbf{E}_1\{|G_{u,c}| \mid M_1 = u\} \\ &= \mathbf{E}_1\left\{|G_{u,c}| \mid \max_{G_{u,c}} S_k = u\right\} \mathbf{P}_1\left\{\max_{G_0} S_k = M_1 \mid M_1 = u\right\} \\ &\quad + \mathbf{E}_1\left\{|G_{u,c}| \mid \max_{G_{u,c}} S_k < u\right\} \mathbf{P}_1\left\{\max_{G_0} S_k < M_1 \mid M_1 = u\right\},\end{aligned} \tag{2.13}$$

where for $u < c$

$$\mathbf{E}_1\left\{|G_{u,c}| \mid \max_{G_{u,c}} S_k = u\right\}$$
$$= \mathbf{E}_1\left\{\tau(u) \mid \tau(u) < \tau(u-c), \ S_{\tau(u)} = u\right\} + \mathbf{E}_1\left\{\tau(-c) \mid \tau_+ = \infty\right\}. \tag{2.14}$$

The following lemma evaluates the asymptotics of the second term of (2.14). The first term is considered below.

Lemma 2.3 *Under the hypothesis of Theorem 2.2,*

$$\mathbf{E}_1\{\tau(-c)|\tau_+ = \infty\} = \rho_1^{-1}\left(c + \frac{\mathbf{E}_1 Z_1^2}{2\rho_1} - 2\,\mathbf{E}_1 M_1\right) - \mathbf{E}_1\tau(M_1) + o(e^{-kc}) \tag{2.15}$$

for some positive k, as $c \to \infty$.

Proof: Consider the queueing process W_k (cf. Feller (1966), vol. II, pp. 193–199) defined as follows,

$$W_0 = 0;$$

$$W_{k+1} = \min\{0;\, W_k - Z_{k+1}\} \ \text{ for } \ k \geq 0.$$

A similar process was discussed in Chapter 1 in the context of CUSUM procedures.

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

$$\mathbf{E}_1\{\tau(-c)|\tau_+ = \infty\} = \mathbf{E}_1\{\tau^W(-c)|\tau_+ = \infty\}. \tag{2.16}$$

It is easy to see that $W_k = S_k^{(1)} - \max\limits_{j \leq k} S_j^{(1)}$. 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

$$\mathbf{E}_1\{\tau^W(-c)|\tau_+ = \infty\} = \mathbf{E}_1\tau^{W^*}(-c) = \mathbf{E}_1\tau^W(-c) - \mathbf{E}_1\tau(M_1).$$

The first passage time for queueing processes, $\tau^W(-c)$, was studied in the literature during the last two decades. Khan (1979b), Lemma 1, proves the existence of the limit

$$\lim_{c\to\infty} \frac{\mathbf{E}_1\tau^W(-c)}{c} = \frac{1}{\rho_1}.$$

Siegmund (1985), formula (2.53), suggests a simple but crude computational formula for $\mathbf{E}_1\tau^W(-c)$ based on Wald approximations,

$$\mathbf{E}_1\tau^W(-c) \approx \frac{c - 1 + e^{-c}}{\rho_1}.$$

The exact asymptotics as $c \to \infty$, up to an exponentially small error term, was obtained by Lotov (1991),

$$\mathbf{E}_1\tau^W(-c) = \frac{c + \beta_1}{\rho_1} + o(e^{-kc}), \quad \text{as } c \to \infty, \tag{2.17}$$

for some $k > 0$, where

$$\beta_1 = \frac{\mathbf{E}_1 Z_1^2}{2\rho_1} - \frac{2\,\mathbf{E}_1\{S_{\tau+}|\tau_+ < \infty\}}{\mathbf{P}_1\{\tau_+ = \infty\}}.$$

Next, we note that $M_1 = \sum_{k=1}^{q} T_k$, where T_k are consecutive ascending ladder heights of the random walk $S^{(1)}$ and q is the number of them. Clearly, q is a geometric random variable with parameter $\mathbf{P}_1\{\tau_+ = \infty\} = \eta_1$, independent of T_k. Also, T_k are i.i.d. random variables whose distribution coincides with a conditional distribution of $S_{\tau+}$ given $\tau_+ = \infty$. Hence, $\mathbf{E}_1 M_1 = \mathbf{E}\, q \cdot \mathbf{E}\, T_1$, and

$$\mathbf{E}_1\{S_{\tau+}|\tau_+ < \infty\} = \mathbf{E}\, T_1 = \frac{\mathbf{E}_1 M_1}{\mathbf{E} q} = \eta_1\, \mathbf{E}_1 M_1.$$

Then $\beta_1 = \mathbf{E}_1 Z_1^2/(2\rho_1) - 2\,\mathbf{E}_1 M_1$. Finally, (2.16) and (2.17) imply (2.15). □

Since $\mathbf{P}_1\{\tau(u) < \tau(u-c) \mid M_1 = u\} \to 1$ as $c \to \infty$, formulae (2.14) and (2.15) yield

$$\begin{aligned}&\mathbf{E}_1\{|G_{u,c}| \mid \max_{G_{u,c}} S_k = u\}\\ &\quad = \mathbf{E}_1\{\tau(u)|\tau(u) < \infty,\ S_{\tau(u)} = u\} + \frac{c+\beta_1}{\rho_1} + \mathbf{E}_1\tau(M) + o(1),\ c \to \infty.\end{aligned} \tag{2.18}$$

Our next step is to show that only the first term is essential in right-hand side of (2.13). We proceed with the following

Lemma 2.4 *For any u from the support of $Q_1(\cdot)$,*

$$-\frac{1}{c}\log\ \mathbf{P}\left\{\max_{G_0} S_k^{(1)} < M_1 \mid M_1 = u\right\} \geq 1 + o(1), \tag{2.19}$$

as $c \to \infty$.

Proof: In terms of the first passage time,

$$\begin{aligned}\mathbf{P}\{\max_{G_0} S_k^{(1)} < M_1 \mid M_1 = u\} &= \mathbf{P}_1\{\tau(u-c) < \tau(u) \mid M_1 = u\}\\ = \mathbf{P}_1\{\tau(u-c) < \tau(u) \mid \tau(u) < \infty\} &= \frac{\mathbf{P}_1\{\tau(u-c) < \tau(u) < \infty\}}{\mathbf{P}_1\{\tau(u) < \infty\}}.\end{aligned}$$

Consider the random walk

$$S_k^* = S_{k+\tau(u-c)}^{(1)} - S_{\tau(u-c)}^{(1)} \text{ for } k > 0. \tag{2.20}$$

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 $S_{\tau(u-c)}^{(1)}$ and has the same distribution as $S_k^{(1)}$. Then 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 $\tau_c^{(1)}$. According to (2.5),

$$\mathbf{P}_1\{\tau(u-c) < \tau(u) < \infty\} \leq \; \mathbf{P}_1\{\tau_c^* < \infty\} \sim e^{-c}\eta_1\eta_2'/\rho_1,$$

as $c \to \infty$, which implies (2.19). $\square$

Using the indicator notation ($I\{A\} = 1$ if the event A occurs and $I\{A\} = 0$ otherwise), one can notice that $\max\limits_{G_{u,c}} S_k^{(1)}$ and

$$|G_{u,c}| = \sum_{k=0}^{\tau(u-c)} I\left\{S_k^{(1)} > u - c\right\}$$

are positively associated random variables (see Lehmann (1966)). Indeed, they are non-decreasing functions of $S_1^{(1)}, S_2^{(1)}, \ldots$, and therefore, are non-decreasing functions of the i.i.d. random variables $\log\frac{f}{g}(X_1)$, $\log\frac{f}{g}(X_2)$, ... This property implies that by (2.18),

$$\mathbf{E}_1\left\{|G_{u,c}| \mid \max_{G_{u,c}} S_k < u\right\} \;\leq\; \mathbf{E}_1\left\{|G_{u,c}| \mid \max_{G_{u,c}} S_k = u\right\} \;=\; O(c).$$

Then the second term of (2.13) tends to 0, because it is at most a product of the linearly increasing and exponentially decreasing functions of c. Also, according to Lemma 2.4,

$$\mathbf{P}\left\{\max_{G_0} S_k^{(1)} = M_1 \mid M_1 = u\right\} \;=\; 1 + O(e^{-c}), \;\text{ as }\; c \to \infty.$$

Hence, from (2.13) and (2.18),

$$\begin{aligned}&\mathbf{E}\left\{|G_0| \mid M_1 = u \geq M_2\right\}\\ &= \mathbf{E}_1\{\tau(u)|\tau(u) < \infty,\ S_{\tau(u)} = u\} + \frac{c+\beta_1}{\rho_1} + \mathbf{E}_1\tau(M) + o(1),\ c \to \infty.\end{aligned}$$

We show now that $\mathbf{E}_1 \sum_{i=1}^{n_1} \{|G_i| \mid M_1 = u\}$ has a finite limit as $c \to \infty$, which does not depend on u. It follows from (2.19) that for large values of c all the moments k defined by $S_k^{(1)} = M_1$ do not belong to $\bigcup\{G_i,\ i \geq 1\}$ with probability tending to 1. Hence the limiting conditional distribution of $\bigcup\{G_i,\ i \geq 1\}$ 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^*, defined by (2.20), and $S_k^{(1)}$ have the same conditional distribution, and

$$\begin{aligned}&\mathbf{E}_1\left\{\sum_{i=1}^{n_1} |G_i| \mid M_1 = u\right\}\\ &= \mathbf{E}_1\left\{\text{number of k satisfying } S_k^* > c + S^{(1)}_{\tau(-c)}| \mid M_1 = u\right\} \to \mathbf{E}|G_+|,\end{aligned}$$

where $G_+ = \{k : S_k^{(1)} > Y\}$ with Y denoting a random variable independent on S_k whose distribution is the limiting distribution of $c + S^{(1)}_{\tau(-c)}$.

Consider now the last term of (2.12). As above,

$$\mathbf{E}_2\left\{\sum_{i=1}^{n_2} |H_i| \mid M_2 \leq u = M_1\right\} \to \mathbf{E}_2|H_+|,$$

where $|H_+|$ is defined similarly to G_+. Thus only the asymptotic distribution of $|H_0|$ depends on c. Lemma 3.2 can be used to evaluate the asymptotics of the expected size of $|H_0|$ if we fix the value of M_2 and condition on it. One has

$$\mathbf{E}_2\{|H_0| \mid M_2 = v \leq M_1 = u\}$$

$$= \mathbf{E}_2\{\tau(v)\,|\,\tau(v) < \tau(u-c),\ S_{\tau(v)} = v\} + \mathbf{E}_2\{\tau(-v+u-c)\,|\,\tau_+ = \infty\} - 1$$
$$= \mathbf{E}_2\{\tau(v)\,|\,\tau(v) < \infty,\ S_{\tau(v)} = v\} + \frac{c+v-u+\beta_2}{\rho_2} - \mathbf{E}_2\tau(M) - 1 + o(1), \tag{2.21}$$

as $c \to \infty$. Here we applied Lemma 3.2 to $S_k^{(2)}$ and $\tau(-v+u-c)$ and defined β_2 similarly to β_1.

Thus (2.12), (2.18) and (2.21) yield

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

where

$$w_1(u,v) = \mathbf{E}_1\{\tau(u)\,|\,\tau(u) < \infty,\ S_{\tau(u)} = u\} + \mathbf{E}_2\{\tau(v)\,|\,\tau(v) < \infty,\ S_{\tau(v)} = v\}$$
$$+\frac{\beta_1}{\rho_1} - \frac{u-v-\beta_2}{\rho_2} - \mathbf{E}_1\tau(M) - \mathbf{E}_2\tau(M) + \mathbf{E}_1|G_+| + \mathbf{E}_2|H_+| - 1.$$

A similar formula holds for $\mathbf{E}\,\{|R_c|\,|\,M_2 = u > M_1 = v\}$, and $w_2(u,v)$ is defined by analogy with $w_1(u,v)$.

Next, we note that

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

$$(ii)\quad \int_{u\geq 0} \mathbf{E}_j\{\tau(u)\,|\,\tau(u) < \infty,\ S_{\tau(u)} = u\}dQ_j(u) = \mathbf{E}_j\tau(M) \quad \text{for } j = 1,2,$$

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

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

where $x^+ = \max\{0, x\}$. Then, from (2.11),

$$\begin{aligned} \mathbf{E}\,|R_c| &= \int_{u\geq 0}\int_{v\leq u} \mathbf{E}\,\{|R_c|\,|\,M_1 = u \geq M_2 = v\}dQ_2(v)dQ_1(u) \\ &\quad + \int_{u\geq 0}\int_{v<u} \mathbf{E}\,\{|R_c|\,|\,M_2 = u > M_1 = v\}dQ_1(v)dQ_2(u) \\ &= c\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) + b + o(1), \end{aligned}$$

as $c \to \infty$, where

$$\begin{aligned} b &= \int_{u\geq 0}\int_{v\leq u} w_1(u,v)\, dQ_2(v) dQ_1(u) + \int_{u\geq 0}\int_{v<u} w_2(u,v) dQ_1(v)\, dQ_2(u) \\ &= \mathbf{E}\left\{\frac{\beta_1 - (M_2 - M_1)^+}{\rho_1} + \frac{\beta_2 - (M_1 - M_2)^+}{\rho_2}\right\} + \mathbf{E}_1|G_+| + \mathbf{E}_2|H_+| - 1 \\ &= \frac{\mathbf{E}_1 Z_1^2/2\rho_1 - \mathbf{E}(M+M_1)}{\rho_1} + \frac{\mathbf{E}_2 Z_1^2/2\rho_2 - \mathbf{E}(M+M_2)}{\rho_2} \\ &\qquad + \mathbf{E}_1|G_+| + \mathbf{E}_2|H_+| - 1. \end{aligned} \tag{2.22}$$

□

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 $c \to \infty$

$$E|R_c| \sim \frac{4c}{\Delta^2} \quad \text{and} \quad \mathbf{P}_\nu\{\nu \in R_c\} \sim \left[1 - \frac{2\exp\{-c-\kappa\}}{\Delta^2}\right]^2,$$

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

Using the Theorems 2.1 and 2.2, one obtains the asymptotic representation for the risk of the confidence set R_c under the loss function (1.1) as $c \to \infty$

$$\begin{aligned} Risk(W, R_c) &= 1 - \left(1 - \frac{e^{-c}\eta_1\eta_2'}{\rho_1}\right)\left(1 - \frac{e^{-c}\eta_1'\eta_2}{\rho_2}\right) + \\ &+ \lambda c\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right) + \lambda(b + o(1)), \end{aligned}$$

where ρ_1 and ρ_2 are defined by (2.2), η_1, η_2, η_1' and η_2' are defined by (2.3).

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

$$\begin{aligned} c &= \log\frac{4\eta_1\eta_2'}{\rho_1+\rho_2} - \log\left(1 - \sqrt{1 - \frac{8\lambda}{\rho_1+\rho_2}}\right) \\ &\sim \log\frac{4\eta_1\eta_2'}{\rho_1+\rho_2} - \log\frac{4\lambda}{\rho_1+\rho_2} = -\log\lambda + \log(\eta_1\eta_2'). \end{aligned} \tag{2.23}$$

With this choice of c,

$$Risk(W, R_{c(\lambda)}) \sim -\lambda\log\lambda\left[\frac{1}{\rho_1} + \frac{1}{\rho_2}\right] + \lambda\left[\frac{(1+\log(\eta_1\eta_2'))(\rho_1+\rho_2)}{\rho_1\rho_2} + b\right] \to 0$$

as $\lambda \to 0$.

Chapter 3

Confidence estimation in exponential families

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 θ_1 and θ_2, $\theta_2 \neq \theta_1$ although our results can be extended to the multivariate case. Function $\psi(\theta)$ is known to be analytic convex function on Θ which is a subinterval of the real line (see for example Brown(1986)). We assume that interval $[\theta_1, \theta_2]$ is contained in $Int\,\Theta$. It is also assumed that

$$0 < L = \liminf_{n\to\infty} \frac{\nu}{n} \leq \limsup_{n\to\infty} \frac{\nu}{n} = U < 1. \tag{3.1}$$

Thus our problem of an interval estimating ν may be considered as the Bayes estimation problem with the discrete uniform prior distribution on $\{Ln, \ldots, Un\}$ and the risk function $\lambda\, \mathbf{E}|R| - \mathbf{P}\{\nu \in R\} + 1$, when θ_1 and θ_2 are nuisance parameters. Assumption (3.1) may seem to be strong, but when the information about F and G is not complete, to detect a change point one needs a sufficient amount of data before and after it. Therefore it is usually assumed that $\nu = \tau n$ for some $0 < \tau < 1$, and τ is to be estimated.

For any fixed $1 \le k \le n$ let $\hat{\theta}_1$ and $\hat{\theta}_2$ be maximum likelihood estimators of θ_1 and θ_2. Then the Bayes estimator of ν with respect to the zero-one loss function and the given prior is

$$\hat{\nu} = \arg \max_{Ln \le k \le Un} \Lambda_k,$$

the overall maximum likelihood estimator, where

$$\Lambda_k = \log \prod_{j=1}^{k} f(X_j|\hat{\theta}_1) \prod_{j=k+1}^{n} f(X_j|\hat{\theta}_2) \,/\, \prod_{j=1}^{n} f(X_j|0). \tag{3.2}$$

By analogy with the previous chapter, define for a positive c the confidence set

$$T_c = \{k : Ln \le k \le Un,\ \Lambda_{\hat{\nu}} - \Lambda_k < c\}.$$

An interesting geometric representation for the set T_c can be given for the case when $\{f(x|\theta)\}$ is a normal distribution family with location parameter θ. We assume that θ is unknown and changes at an unknown time point, but the standard deviation σ is known and fixed. In that case for $\tilde{\Lambda}_k = 2\sigma^2 \left\{\Lambda_k + \log(\sigma\sqrt{2\pi})\right\} + \sum_1^n X_j^2$ one has

$$\tilde{\Lambda}_k = \frac{1}{k}\left(\sum_{j=1}^{k} X_j\right)^2 + \frac{1}{n-k}\left(\sum_{j=k+1}^{n} X_j\right)^2.$$

Let $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbf{R}^n$ be the vector of observations, and for each $Ln \le k \le Un$ introduce a pair of n-dimensional vectors

$$\mathbf{a}_k = \left(\underbrace{\frac{1}{\sqrt{k}}, \ldots, \frac{1}{\sqrt{k}}}_{k}, 0, \ldots, 0\right) \text{ and } \mathbf{b}_k = \left(0, \ldots, 0, \underbrace{\frac{1}{\sqrt{n-k}}, \ldots, \frac{1}{\sqrt{n-k}}}_{n-k}\right).$$

Clearly, $\|\mathbf{a}_k\| = \|\mathbf{b}_k\| = 1$ and $\mathbf{a}_k \perp \mathbf{b}_k$. Then for any k, with $<,>$ denoting inner product,

$$\tilde{\Lambda}_k = (<\mathbf{a}_k, \mathbf{X}>)^2 + (<\mathbf{b}_k, \mathbf{X}>)^2 = [\ \Pr{}_{V_k}(\mathbf{X})]^2,$$

is the squared orthogonal projection of the vector of data $\mathbf{X}$ onto the plane V_k, formed by $\mathbf{a}_k$ and $\mathbf{b}_k$. Thus, in order to maximize Λ_k, which is the same as to maximize $\tilde{\Lambda}_k$, the set T_c selects such planes from $V_{Ln}, \ldots, V_{Un}$ that are the closest to $\mathbf{X}$ in a sense of the largest projection of $\mathbf{X}$ onto V_k, or the smallest angle between them.

Confidence estimation of the change point in the presence of nuisance parameters was studied in Siegmund (1988) and James, James and Siegmund (1992) for the cases of univariate and multivariate normal distributions respectively. Our main goal is to study the behavior of T_c in the general case of an arbitrary exponential family in terms of the given risk function and to suggest the optimal choice of c. The formulae for the asymptotic risk of T_c as $n \to \infty$ follow from the proximity of set T_c and set R_c, which was defined by (2.1) when θ_1 and θ_2 were supposed to be known. To establish this proximity we proceed with the following two propositions.

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

$$1 - \mathbf{P}\left\{T_c \subset [\nu - n^\alpha, \nu + n^\alpha]\right\} = o\left(e^{-Kn^\alpha}\right)$$

as $n \to \infty$ *for some positive constant* K.

Proof:

Let $H(x) = \sup_\theta \{\theta x - \psi(\theta)\}$ be the Legendre transform of the function ψ. Then, from (3.2) and the definition of $\hat{\theta}_1$ and $\hat{\theta}_2$,

$$\begin{aligned} \Lambda_k &= k(\hat{\theta}_1 \bar{x}_k - \psi(\hat{\theta}_1)) + (n-k)(\hat{\theta}_2 \bar{x}_{kn} - \psi(\hat{\theta}_2)) \\ &= kH(\bar{x}_k) + (n-k)H(\bar{x}_{kn}). \end{aligned} \qquad (3.3)$$

Here and later we use the notation

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

The maximum likelihood estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, which maximize the function $\theta x - \psi(\theta)$ for $x = \bar{x}_k$ and $x = \bar{x}_{kn}$, are the roots of the equation $\psi'(\theta) = x$ for these values of x. Since ψ' is an increasing function of θ, 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 $\theta_2 = h(\mu_2)$ with μ_j denoting $\mathbf{E}_{\theta_j} X$.

We proceed with the following lemma which will be used a number of times in this chapter.

Lemma 3.2 *Let $\xi_1, \xi_2, \ldots, \xi_k$ be i.i.d. random variables with density $f(x|\theta)$ for some $\theta \in$ Int Θ with $\mu = \mathbf{E}\,\xi_1$. Let Δ be a closed subset of Θ containing θ in its interior, and*

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

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

$$\mathbf{P}\left\{|\bar{\xi} - \mu| \geq \varepsilon\right\} \ \leq \ 2e^{-m\varepsilon^2 k/2}.$$

Proof: Since $f(x|\theta)$ is an exponential family density, its moment generating function

$$\mathbf{E}_\theta e^{\lambda\xi_1} = \exp\{\psi(\theta + \lambda) - \psi(\theta)\} \tag{3.4}$$

exists at least for $|\lambda|$ smaller than the distance from θ to $\Theta^c = \mathbf{R}\backslash\Theta$. Hence by the classical results from the theory of large deviations (see e.g. Bahadur (1971))

$$\mathbf{P}\left\{|\bar{\xi} - \mu| \geq \varepsilon\right\} \ \leq \ 2\exp\{-k \min_{|t|=\varepsilon} G(\mu + t)\}, \tag{3.5}$$

where $G(a) = \sup_\lambda (a\lambda - \log\ \mathbf{E}e^{\lambda\xi_1})$. Then, from (3.4),

$$G(a) = \sup_\lambda \{a(\theta + \lambda) - \psi(\theta + \lambda)\} - a\theta + \psi(\theta) = H(a) - a\theta + \psi(\theta).$$

Then for the function $\phi(t) = H(\mu + t) - (\mu + t)\theta + \psi(\theta)$ we have

$$\phi(0) = H(\mu) - \mu\theta + \psi(\theta) = 0;$$
$$\phi'(0) = h(\mu) - \theta = 0;$$
$$\phi''(t) = h'(\mu + t) = \frac{1}{\psi''(h(\mu + t))} > 0.$$

Hence, for $t = \pm\varepsilon$ and some $-\varepsilon \le t_1 \le \varepsilon$

$$\phi(t) = t^2\phi''(t_1)/2 \ge \varepsilon^2 \min_{|t|\le\varepsilon} \phi''(t)/2 = \varepsilon^2 m/2,$$

and Lemma 3.2 follows from (3.5). □

To apply this Lemma to our situation let Δ be any closed subinterval of Θ which contains θ_1 and θ_2 in its interior and let $m = \min\{1/\psi''(t), t \in \Delta\} = \min\{h'(x), x \in \psi'(\Delta)\}$. Also introduce $M = \max\{1/\psi''(t), t \in \Delta\} = \max\{h'(x), x \in \psi'(\Delta)\} < \infty$.

We continue the proof of Proposition 3.1. Consider only the case $Ln \le k \le \nu - n^\alpha$ (which is trivial if $Ln > \nu - n^\alpha$). Let us show first that the order of $\Lambda_\nu - \Lambda_k$ is at least n^α as $n \to \infty$ with the probability tending to 1.

Using (3.3) one obtains

$$\begin{aligned}\Lambda_\nu - \Lambda_k &= \nu H(\bar{x}_\nu) - kH(\bar{x}_k) + (n-\nu)H(\bar{x}_{\nu n}) - (n-k)H(\bar{x}_{kn}) \\ &= (\nu - k)H(\mu_1) + (n-\nu)H(\mu_2) - (n-k)H(\tilde{\mu}) + r_k, \qquad (3.6)\end{aligned}$$

where $\tilde{\mu} = \mathbf{E}\,\bar{x}_{kn} = \dfrac{\nu-k}{n-k}\mu_1 + \dfrac{n-\nu}{n-k}\mu_2$.

Let $g(t) = H(t\mu_1 + (1-t)\mu_2)$, which is clearly a convex function. Then with $t = (\nu-k)/(n-k)$ the formula (3.6) can be rewritten as

$$\Lambda_\nu - \Lambda_k = (n-k)\{tg(1) + (1-t)g(0) - g(t)\} + r_k.$$

According to assumption (3.1), for any k $t = (\nu-k)/(n-k) \le \nu/n \le U$. For any $0 \le t \le U$

$$g(t) \le \left(1 - \frac{t}{U}\right) g(0) + \frac{t}{U} g(U).$$

Hence

$$\begin{aligned}\Lambda_\nu - \Lambda_k &\ge (n-k)\left\{\left(\frac{t}{U} - t\right) g(0) + tg(1) - \frac{t}{U} g(U)\right\} + r_k \\ &= (n-k)\frac{t}{U}\{(1-U)g(0) + Ug(1) - g(U)\} + r_k,\end{aligned}$$

and for $t = (\nu - k)/(n - k)$,

$$\Lambda_\nu - \Lambda_k \geq K_1(\nu - k) + r_k, \tag{3.7}$$

where $K_1 = [(1-U)g(0) + Ug(1) - g(U)]/U$ is a positive constant.

Consider the remainder r_k. Using the fact that $H'(x) = h(x)$ one derives from (3.6)

$$\begin{aligned} r_k &= \nu\{H(\bar{x}_\nu) - H(\mu_1)\} - k\{H(\bar{x}_k) - H(\mu_1)\} \\ &\quad +(n-\nu)\{H(\bar{x}_{\nu n}) - H(\mu_2)\} - (n-k)\{H(\bar{x}_{kn}) - H(\mu_2)\} \\ &= \nu h(\mu_1)(\bar{x}_\nu - \mu_1) - kh(\mu_1)(\bar{x}_k - \mu_1) \\ &\quad +(n-\nu)h(\mu_2)(\bar{x}_{\nu n} - \mu_2) - (n-k)h(\tilde{\mu})(\bar{x}_{kn} - \tilde{\mu}) + r_k^{(2)}. \end{aligned} \tag{3.8}$$

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

As we observed before, $h(\mu_j) = \theta_j$. Also, $\bar{x}_\nu = \dfrac{\nu - k}{\nu}\bar{x}_{k\nu} + \dfrac{k}{\nu}\bar{x}_k$ and $\bar{x}_{kn} = \dfrac{\nu - k}{n-k}\bar{x}_{kn} + \dfrac{n-\nu}{n-k}\bar{x}_{\nu n}$. Therefore

$$\begin{aligned} r_k &= (\nu - k)\theta_1(\bar{x}_{k\nu} - \mu_1) + (n-\nu)\theta_2(\bar{x}_{\nu n} - \mu_2) \\ &\quad - h(\tilde{\mu})[(\nu-k)(\bar{x}_{k\nu} - \mu_1) + (n-\nu)(\bar{x}_{\nu n} - \mu_2)] + r_k^{(2)} \\ &= (\nu-k)(\theta_1 - h(\tilde{\mu}))(\bar{x}_{k\nu} - \mu_1) + (n-\nu)(\theta_2 - h(\tilde{\mu}))(\bar{x}_{\nu n} - \mu_2) + r_k^{(2)}. \end{aligned} \tag{3.9}$$

In fact the second term of (3.9) has also the order $\nu - k$ because the Taylor expansion of $h(\tilde{\mu})$ shows that

$$\begin{aligned} |(n-\nu)(\theta_2 - h(\tilde{\mu}))| = |(n-\nu)h'(\zeta_1)(\mu_2 - \tilde{\mu})| = (\nu - k)\frac{n-\nu}{n-k}|\mu_1 - \mu_2|h'(\zeta_1) \\ \leq (\nu - k)|\mu_1 - \mu_2| \max_{\zeta_1 \in \psi'(\Delta)} h'(\zeta_1) = (\nu - k)M|\mu_1 - \mu_2| \end{aligned}$$

with ζ_1 between μ_1 and μ_2. Since $h(x)$ is an increasing function, $h(\zeta_1) \in \Delta$.

Similarly $|\theta_1 - h(\tilde{\mu})| \leq M|\mu_1 - \mu_2|$ and (3.9) shows that

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

Using Lemma 3.2 with $\varepsilon = \dfrac{K_1}{6M|\mu_1-\mu_2|} = K_2$ one obtains for sufficiently large n

$$\mathbf{P}\left\{|r_k^{(1)}| \geq \frac{K_1}{3}(\nu-k)\right\} \leq \mathbf{P}\left\{M|\mu_1-\mu_2|(|\bar{x}_{k\nu}-\mu_1|+|\bar{x}_{\nu n}-\mu_2|) \geq \frac{K_1}{3}\right\}$$

$$\leq \mathbf{P}\left\{\max(|\bar{x}_{k\nu}-\mu_1|;|\bar{x}_{\nu n}-\mu_2|) \geq \frac{K_1/2}{3M|\mu_1-\mu_2|}\right\}$$

$$\leq 4\exp\left\{-\frac{mK_2^2}{2}\min(\nu-k,n-\nu)\right\} \leq 4\exp\{-\frac{mK_2^2}{2}n^\alpha\}. \tag{3.10}$$

The error term of the Taylor expansion of (3.9) also can be estimated using Lemma 3.2:

$$\begin{aligned} |r_k^{(2)}| &= \frac{1}{2}\,|\,\nu h'(\zeta_2)(\bar{x}_\nu-\mu_1)^2 - kh'(\zeta_3)(\bar{x}_k-\mu_1)^2 \\ &\quad +(n-\nu)h'(\zeta_4)(\bar{x}_{\nu n}-\mu_2)^2 - (n-k)h'(\zeta_5)(\bar{x}_{kn}-\tilde{\mu})^2\,|\,, \end{aligned}$$

where ζ_2, ζ_3, ζ_4 and ζ_5 are some 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 δ_j the distance from μ_j to the boundary of $\psi'(\Delta)$ for $j = 1, 2$ and let $\delta = \min(\delta_1, \delta_2)$. Then

$$\mathbf{P}\{\zeta_2 \notin \psi'(\Delta)\} \leq \mathbf{P}\{|\zeta_2-\mu_1| \geq \delta\} \tag{3.11}$$

$$\leq \mathbf{P}\{|\bar{x}_\nu-\mu_1| \geq \delta\} \leq 2\exp\{-m\delta^2\nu/2\}.$$

Probabilities involving ζ_3, ζ_4 and ζ_5 can be estimated similarly. So, if $\gamma = \min(\nu, k, n-\nu, n-k) = \min(k, n-\nu) \geq n\min(L, 1-U)$, then

$$|r_k^{(2)}| \leq \frac{M}{2}|\nu(\bar{x}_\nu-\mu_1)^2 - k(\bar{x}_k-\mu_1)^2 + (n-\nu)(\bar{x}_{\nu n}-\mu_2)^2 - (n-k)(\bar{x}_{kn}-\tilde{\mu})^2| \tag{3.12}$$

with probability greater than $1-8\exp\{-m\delta^2\gamma/2\}$. Applying Lemma 3.2 once again we get

$$\mathbf{P}\left\{\nu(\bar{x}_\nu-\mu_1)^2 > \varepsilon\right\} = \mathbf{P}\left\{|\bar{x}_\nu-\mu_1| \geq \sqrt{\frac{\varepsilon}{\nu}}\right\} \leq 2e^{-m\varepsilon/2}.$$

The same upper bound can be obtained for the other terms of (3.12). Thus Lemma 3.2 with $\varepsilon = K_1/(6M) = K_3$ and (3.12) yield

$$\mathbf{P}\left\{|r_k^{(2)}| \geq \frac{K_1}{3L}(\nu - k)\right\} \leq 8[\exp\{-m\delta^2\gamma/2\} + \exp\{-mK_3 n^\alpha/2\}].$$

By combining this result with (3.10) one obtains

$$\mathbf{P}\left\{|r_k| \geq \frac{2}{3}K_1(\nu - k)\right\} \leq$$
$$12\exp\{-K_4 n^\alpha\} + o\left(\exp\{-K_4 n^\alpha\}\right) \tag{3.13}$$

as $n \to \infty$, where $K_4 = \min(K_2, K_3)/2$. This estimate and (3.7) show that with probability (3.13)

$$\Lambda_\nu - \Lambda_k \geq \frac{K_1}{3}(\nu - k) \geq \frac{K_1}{3}n^\alpha. \tag{3.14}$$

By the definition of $\hat{\nu}$ event (3.14) implies that $\Lambda_{\hat{\nu}} - \Lambda_k \geq K_1 n^\alpha/3$, which shows that $k \notin T_c$ for any $n > (3c/K_1)^{1/\alpha}$. Thus as $n \to \infty$ for any $K < K_4$

$$\mathbf{P}\left\{[Ln, \nu - n^\alpha]\bigcap T_c = \emptyset\right\} \leq \nu\left[12e^{-K_4 n^\alpha} + o\left(e^{-K_4 n^\alpha}\right)\right]$$
$$+ n\left[12Ue^{-K_4 n^\alpha} + o\left(e^{-K_4 n^\alpha}\right)\right] = o\left(e^{-Kn^\alpha}\right). \tag{3.15}$$

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

$$\mathbf{P}\left\{[\nu + n^\alpha, Un]\bigcap T_c = \emptyset\right\} = o\left(e^{-K'n^\alpha}\right) \quad \text{as} \quad n \to \infty$$

with some positive constant K'. This result together with (3.15) proves the Proposition 3.1.

□

Proposition 3.3 *Let $\varepsilon_n = rn^{-\beta}$ for some $0 \leq \beta < 1/2 - \alpha$ and $r > 0$. Then*

$$\mathbf{P}\left\{\max_{[\nu - n^\alpha, \nu + n^\alpha]} |(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)| \geq \varepsilon_n\right\} = o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\}\right)$$

as $n \to \infty$ for some $C > 0$.

Proof:

Assume that $\nu - n^\alpha \le k \le \nu$ and consider function $q(\theta, x) = \log \dfrac{f(x|\theta)}{f(x|0)} = \theta x - \psi(\theta)$ for which $H(x) = q(h(x), x)$. By the definition of S_k

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

and (3.6) implies

$$\begin{aligned}
&(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k) \\
&\quad = \nu[H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu)] - k[H(\bar{x}_k) - q(\theta_1, \bar{x}_k)] \\
&\qquad + (n-\nu)[H(\bar{x}_{\nu n}) - q(\theta_2, \bar{x}_{\nu n})] - (n-k)[H(\bar{x}_{kn}) - q(\theta_2, \bar{x}_{kn})] \\
&\quad = A_1 + A_2 + A_3, \qquad (3.16)
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= (\nu - k)\left\{[H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu)] - [H(\bar{x}_{\nu n}) - q(\theta_2, \bar{x}_{\nu n})]\right\}; \\
A_2 &= k\left\{H(\bar{x}_\nu) - H(\bar{x}_k) - [q(\theta_1, \bar{x}_\nu) - q(\theta_1, \bar{x}_k)]\right\}; \\
A_3 &= (n-k)\left\{H(\bar{x}_{\nu n}) - H(\bar{x}_{kn}) - [q(\theta_2, \bar{x}_{\nu n}) - q(\theta_2, \bar{x}_{kn})]\right\}.
\end{aligned}$$

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

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

for some ζ_6 between θ_1 and $h(\bar{x}_\nu)$. Here $\dfrac{\partial q}{\partial \theta}(\theta_1, \bar{x}_\nu) = \bar{x}_\nu - \psi'(\theta_1) = \bar{x}_\nu - \mu_1$ and $\dfrac{\partial^2 q}{\partial \theta^2}(\zeta_6, \bar{x}_\nu) = -\psi''(\zeta_6)$. The Taylor formula for $h(x)$ shows that

$$\begin{aligned}
H(\bar{x}_\nu) - q(\theta_1, \bar{x}_\nu) &= h'(\zeta_7)(\bar{x}_\nu - \mu_1)^2 - \psi''(\zeta_6)\left\{h'(\zeta_7)(\bar{x}_\nu - \mu_1)\right\}^2/2 \\
&= h'(\zeta_7)\left(1 - \psi''(\zeta_6)h'(\zeta_7)/2\right)(\bar{x}_\nu - \mu_1)^2,
\end{aligned}$$

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

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

$$\begin{aligned} |A_1| &\le (\nu - k) M \left(1 + \frac{M}{2m}\right) \left\{(\bar{x}_\nu - \mu_1)^2 + (\bar{x}_{\nu n} - \mu_2)^2\right\} \\ &\le 2n^\alpha M \left(1 + \frac{M}{2m}\right) \max\left\{(\bar{x}_\nu - \mu_1)^2, (\bar{x}_{\nu n} - \mu_2)^2\right\}. \end{aligned}$$

Hence by Lemma 3.2

$$\begin{aligned} \mathbf{P}\left\{|A_1| \ge \frac{\varepsilon_n}{3}\right\} \le \mathbf{P}&\left\{\max\left\{|\bar{x}_\nu - \mu_1|, |\bar{x}_{\nu n} - \mu_2|\right\} \ge \sqrt{\frac{\varepsilon_n/3}{2n^\alpha M \left(1 + \frac{M}{2m}\right)}}\right\} \\ &= o\left(\exp\left\{-Cn^{1-\alpha-\beta}\right\}\right) \end{aligned} \tag{3.17}$$

as $n \to \infty$ for some $C > 0$.

The second term of (3.16) can be estimated as follows

$$\begin{aligned} A_2 &= k\left\{H(\bar{x}_\nu) - H(\bar{x}_k) - \theta_1(\bar{x}_\nu - \bar{x}_k)\right\} \\ &= k\left\{h(\bar{x}_k)(\bar{x}_\nu - \bar{x}_k) + h'(\zeta_8)(\bar{x}_\nu - \bar{x}_k)^2/2 - \theta_1(\bar{x}_\nu - \bar{x}_k)\right\} \\ &= k(\bar{x}_\nu - \bar{x}_k)\left\{h(\bar{x}_k) - h(\mu_1) + h'(\zeta_8)(\bar{x}_\nu - \bar{x}_k)/2\right\} \\ &= k(\bar{x}_\nu - \bar{x}_k)\left\{h'(\zeta_9)(\bar{x}_k - \mu_1) + h'(\zeta_8)(\bar{x}_\nu - \bar{x}_k)/2\right\}. \end{aligned}$$

As in (3.11), the intermediate values ζ_8 and ζ_9 belong to $\psi'(\Delta)$ with probability $1 - o\left(e^{-Cn}\right)$ as $n \to \infty$, and therefore, since $\bar{x}_\nu - \bar{x}_k = \dfrac{\nu - k}{\nu}(\bar{x}_{k\nu} - \bar{x}_k)$, $k \le \nu$ and $\nu - k \le n^\alpha$,

$$|A_2| \le n^\alpha |\bar{x}_{k\nu} - \bar{x}_k|\, M \max\left\{\frac{1}{2}|\bar{x}_\nu - \bar{x}_k|, |\bar{x}_k - \mu_1|\right\}$$

with the probability that is also $1 - o\left(e^{-Cn}\right)$.

Let $0 < \sigma < 1 - \alpha - \beta$. By Lemma 3.2

$$\mathbf{P}\left\{\max\left\{|\bar{x}_{k\nu} - \bar{x}_k|, |\bar{x}_k - \mu_1|\right\} \ge \frac{\varepsilon_n}{3Mn^{\alpha+\sigma}}\right\} = o\left(\exp\{-Cn^{1-2(\alpha+\beta+\sigma)}\}\right)$$

as $n \to \infty$ for some $C > 0$. Also

$$\mathbf{P}\{|\bar{x}_{k\nu} - \bar{x}_k| \geq n^\sigma\} \leq \mathbf{P}\{|\bar{x}_{k\nu}| \geq n^\sigma/2\} + \mathbf{P}\{|\bar{x}_k| \geq n^\sigma/2\} \leq 2P_{\theta_1}\{|X_1| \geq n^\sigma/2\},$$

which is $o(\exp\{-Cn^\sigma\})$ for some C. Indeed, by (3.4) for some small λ

$$\Psi(\lambda) = \mathbf{E}_{\theta_1} e^{\lambda|X_1|} \leq \mathbf{E}_{\theta_1} e^{\lambda X_1} + \mathbf{E}_{\theta_1} e^{-\lambda X_1} = \frac{e^{\psi(\theta_1+\lambda)} + e^{\psi(\theta_1-\lambda)}}{e^{\psi(\theta_1)}} < \infty,$$

and

$$\mathbf{P}_{\theta_1}\{|X_1| \geq n^\sigma/2\} = \int_{|x|\geq n^\sigma/2} f(x|\theta_1)d\mu(x)$$
$$\leq \exp\{-\lambda n^\sigma/2\} \int_{|x|\geq n^\sigma/2} e^{\lambda|x|} f(x|\theta_1)d\mu(x) \leq \Psi(\lambda)\exp\{-\lambda n^\sigma/2\}.$$

Hence, for any $0 < \sigma < 1/2 - \alpha - \beta$,

$$\mathbf{P}\{|A_2| \geq \varepsilon_n/3\} = o\left(\exp\left\{-Cn^{1-2(\alpha+\beta+\sigma)}\right\} + \exp\{-Cn^\sigma\}\right).$$

We choose $\sigma = (1 - 2\alpha - 2\beta)/3$ in which case with some C

$$\mathbf{P}\{|A_2| \geq \varepsilon_n/3\} = o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\}\right).$$

The same estimate holds for $\mathbf{P}\{|A_3| \geq \varepsilon_n/3\}$. Finally, by (3.16) and (3.17)

$$\mathbf{P}\{|(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)| \geq \varepsilon_n\} = o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\}\right).$$

The case $\nu < k < \nu + n^\alpha$ can be treated similarly, and one has

$$\mathbf{P}\left\{\max_{[\nu-n^\alpha,\nu+n^\alpha]} |(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)| \geq \varepsilon_n\right\}$$
$$\leq \sum_{k=\nu-n^\alpha}^{\nu+n^\alpha} \mathbf{P}\{|(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)| \geq \varepsilon_n\} = o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\}\right).$$

$\square$

As was observed, $R_c = \{S_{\hat{\nu}} - S_k \leq c\} \subset [\nu - n^\alpha, \nu + n^\alpha]$ with probability tending to one. As will be shown below, the rate of convergence here is also exponential in n^α. Therefore Propositions 3.1 and 3.3 imply the following result.

Theorem 3.4 *Let $\varepsilon_n = rn^{-\beta}$ for some $0 \le \beta < 1/2$ and $r > 0$. Then*

$$\mathbf{P}\left\{R_{c-\varepsilon_n} \subset T_c \subset R_{c+\varepsilon_n}\right\} \to 1 \text{ as } n \to \infty.$$

The rate of convergence is $o\left(\exp\left\{-Kn^{(1-2\beta)/5}\right\}\right)$ for some $K > 0$. Here c can be any bounded positive function of n.

Proof:

Let α be an arbitrary number from interval $(0, 1/2-\beta)$. Then for any $Ln \le k \le \nu - n^\alpha$

$$\begin{aligned} S_\nu - S_k &= \sum_{j=k+1}^{\nu} \log \frac{f(x_j|\theta_1)}{f(x_j|\theta_2)} \\ &= (\nu - k)\left\{(\theta_1 - \theta_2)\bar{x}_{k\nu} - (\psi(\theta_1) - \psi(\theta_2))\right\} \\ &= (\nu - k)\left\{\rho_1 + (\theta_1 - \theta_2)(\bar{x}_{k\nu} - \mu_1)\right\}, \end{aligned}$$

where, clearly,

$$\rho_1 = \mathbf{E}_{\theta_1} \log \frac{f(x_j|\theta_1)}{f(x_j|\theta_2)} = (\theta_1 - \theta_2)\mu_1 - (\psi(\theta_1) - \psi(\theta_2)).$$

It follows from Lemma 3.2 that

$$\mathbf{P}\left\{|\bar{x}_{k\nu} - \mu_1| \ge \frac{\rho_1}{2|\theta_1 - \theta_2|}\right\} = o\left(e^{-Cn^\alpha}\right)$$

for some positive C. Hence

$$\mathbf{P}\left(S_\nu - S_k \ge \frac{\rho_1 n^\alpha}{2}\right) = o\left(e^{-Cn^\alpha}\right),$$

and therefore, $S_{\hat{\nu}} - S_k \ge S_\nu - S_k \ge \dfrac{\rho_1 n^\alpha}{2} > c$ with probability $o\left(e^{-Cn^\alpha}\right)$ for any $n > (2c/\rho_1)^{1/\alpha}$, which implies $k \notin R_c$. The case $\nu + n^\alpha \le k \le Un$ is similar. Thus

$$\mathbf{P}\left\{R_c \not\subset [\nu - n^\alpha, \nu + n^\alpha]\right\}$$

$$= \mathbf{P}(\min\left\{S_\nu - S_k\,, Ln \le k \le Un,\ k \notin [\nu - n^\alpha, \nu + n^\alpha]\right\} < c\} = o\left(e^{-Cn^\alpha}\right). \quad (3.18)$$

Here c is any bounded function of n, so that it can be replaced by $c - \varepsilon_n$. By (3.18) and Proposition 3.1, $R_{c-\varepsilon_n}$ and T_c are both subsets of $[\nu - n^\alpha, \nu + n^\alpha]$ with probability $1 - o\left(e^{-Cn^\alpha}\right)$ for some $C > 0$. Also, by Proposition 3.3,

$$\mathbf{P}\left\{\max_{\nu-n^\alpha \le k \le \nu+n^\alpha} |(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)| \ge \varepsilon/2\right\} = o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\}\right).$$

For any k in $R_{c-\varepsilon_n}$, $S_{\hat\nu} - S_k \le c - \varepsilon_n$ and

$$\begin{aligned}\Lambda_{\hat\nu} - \Lambda_k &= \{(\Lambda_\nu - \Lambda_k) - (S_\nu - S_k)\} + \{(\Lambda_{\hat\nu} - \Lambda_\nu) - (S_{\hat\nu} - S_\nu)\} + (S_{\hat\nu} - S_k) \\ &\le \varepsilon_n/2 + \varepsilon_n/2 + (c - \varepsilon_n) = c \qquad (3.19)\end{aligned}$$

with probability $1 - o\left(\exp\left\{-Cn^{(1-2\alpha-2\beta)/3}\right\} + \exp\left\{-Cn^\alpha\right\}\right)$. But $\Lambda_{\hat\nu} - \Lambda_k \le c$ means that $k \in T_c$. So $R_{c-\varepsilon_n} \subset T_c$ with the probability above for any $0 < \alpha < 1/2 - \beta$. With the optimal choice of $\alpha = (1-2\beta)/5$

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

The other inclusion, $T_c \subset R_{c+\varepsilon_n}$, can be proven similarly by using (3.19).

□

Theorem 3.4 states that the sets T_c and R_c are very close when n is large, so that

$$\lim_{n\to\infty} \mathbf{P}\{\nu \in R_{c-\varepsilon_n}\} \le \lim_{n\to\infty} \mathbf{P}\{\nu \in T_c\} \le \lim_{n\to\infty} \mathbf{P}\{\nu \in R_{c+\varepsilon_n}\}$$

and

$$\lim_{n\to\infty} \mathbf{E}\,|R_{c-\varepsilon_n}| \le \lim_{n\to\infty} \mathbf{E}\,|T_c| \le \lim_{n\to\infty} \mathbf{E}\,|R_{c+\varepsilon_n}|$$

with ε_n as chosen in this Theorem. Setting $n \to \infty$ in both inequalities and using Theorems 2.1 and 3.1 one derives the following corollaries.

Corollary 3.5 *As* $c \to +\infty$

$$\lim_{n\to\infty} (1 - \mathbf{P}\{\nu \in T_c\}) \sim e^{-c}\eta_1\eta_2'\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right),$$

where η_j *and* η_j' *are given by (2.3),* ρ_1 *and* ρ_2 *are information numbers, specified in (2.2).*

In terms of ψ, $\rho_1 = (\theta_2 - \theta_1)\psi'(\theta_2) - (\psi(\theta_2) - \psi(\theta_1))$ and ρ_2 has the similar expression.

Corollary 3.6 *As* $c \to +\infty$

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

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

$$c = \log\frac{4\eta_1\eta_2}{\mu_1 + \mu_2} - \log\left(1 - \sqrt{1 - \frac{8\lambda}{\mu_1 + \mu_2}}\right) \sim -\log\lambda + \log(\eta_2\eta_2).$$

The next result evaluates the limiting probability of the correct decision, so it concerns the point estimation of the change-point parameter. It is included since it is a straightforward corollary of Theorem 3.4.

Corollary 3.7 *If distributions* F *and* G *do not have atoms then*

$$\lim_{n\to\infty} \mathbf{P}\{\nu = \hat{\nu}\} = \mathbf{P}\{\tau_+ = \infty\}\,\mathbf{P}\{\tau_+' = \infty\}, \tag{3.20}$$

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

Proof: Note that because of the absence of atoms $\nu = \hat{\nu}$ is μ-almost surely equivalent to $\Lambda_\nu = \Lambda_{\hat{\nu}}$ or to the event $\nu \in T_0$. Let $\varepsilon_n = n^{-\beta}$ for some $0 \leq \beta < 1$ so that $T_0 = \bigcap T_{\varepsilon_n}$.

According to Theorem 3.4 $R_0 \subset T_{\varepsilon_n} \subset R_{2\varepsilon_n}$ with probability tending to one. Hence as $n \to \infty$ the limiting probability $\mathbf{P}\{\nu = \hat{\nu}\}$ is bounded from below by $\mathbf{P}\{\nu \in R_0\} = \mathbf{P}\{S_{\max} = S_\nu\} = \mathbf{P}\{\tau_+ = \infty\}\,\mathbf{P}\{\tau_+' = \infty\}$ and from above by $\lim_{n\to\infty} \mathbf{P}\{\nu \in R_{2\varepsilon_n}\} = \mathbf{P}\{\nu \in R_0\}$. Thus (3.20) holds. □

Corollary 3.7 was obtained earlier by Hu and Rukhin (1995).

Chapter 4

Convergence rates of the confidence set characteristics

In the previous chapters we studied asymptotic behavior of confidence sets as the sample size increases unboundedly. How accurate are the obtained limiting formulae if in fact n is finite? In other words, can one use the results of Chapters 2 and 3 for practical applications? We answer these questions by studying the convergence rates of the expected width and the coverage probability as $n \to \infty$ and by showing that the error of such an approximation converges to zero exponentially fast. It turns out, that the problems of this kind are closely related to basic concepts of renewal theory. That is why first we need some auxiliary renewal theoretic results.

4.1 Renewal theory: more about the first passage time process

In this section we consider a random walk $S_k = \sum_{j=1}^{k} Z_j$ generated by a sequence of i.i.d. random variables $Z_1, Z_2, \ldots$ with mean $-\infty \leq \mathbf{E}\, Z_1 = \mu < 0$. Define the first passage time *beyond* the level x as

$$t_x = \begin{cases} \inf\{k : S_k > x\} \text{ for } x > 0; \\ \inf\{k : S_k < x\} \text{ for } x < 0; \end{cases}$$

and the first (strong) ascending ladder epoch

$$\tau_+ = \inf\{k : S_k > 0\}.$$

Clearly, $t_x = \min\{\tau(y),\ y > x\}$, where $\tau(y)$ is the first passage time process, which we studied in the proofs of Theorem 2.1 and Theorem 2.2. Here we study some aspects of t_x and τ_+ concerning the asymptotics of their tail probabilities. In general case the exact asymptotic formulae for them are intractable, but one can obtain some reasonable bounds. Let

$$\sigma = \inf_s \ \mathbf{E}\ \exp\{Z_1 s\},$$
$$\eta = \exp\left\{-\sum_1^\infty k^{-1}\ \mathbf{P}\{S_k > 0\}\right\},$$
$$\eta' = \exp\left\{-\sum_1^\infty k^{-1}\ \mathbf{P}\{S_k \geq 0\}\right\}.$$

Lemma 4.1 *For any* $n > \max\{0;\ \frac{\log(1-\sigma)}{\log\sigma} - 1\}$

$$\mathbf{P}\{n < \tau_+ < \infty\} \leq \frac{\sigma^{n+1}\,\eta}{1 - \sigma - \sigma^{n+1}}. \tag{4.1}$$

Proof: One has for the random walk S_k

$$\mathbf{P}\{n < \tau_+ < \infty\} \;=\; \mathbf{P}\left\{(\bigcap_{k=1}^{n}\{S_k \leq 0\}) \cap (\bigcup_{k=n+1}^{\infty}\{S_k > 0\})\right\}$$
$$= \mathbf{P}\left\{\bigcap_{k=1}^{n}\{S_k \leq 0\}\right\}\ \mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\} \mid \bigcap_{k=1}^{n}\{S_k \leq 0\}\right\}$$
$$= (\ \mathbf{P}\{n < \tau_+ < \infty\} +\ \mathbf{P}\{\tau_+ = \infty\})\ \mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\} \mid \bigcap_{k=1}^{n}\{S_k \leq 0\}\right\}.$$

Hence,

$$\mathbf{P}\{n < \tau_+ < \infty\} \;=\; \frac{\eta\;\mathbf{P}\left\{\bigcup\limits_{k=n+1}^{\infty}\{S_k > 0\}\,|\,\bigcap\limits_{k=1}^{n}\{S_k \le 0\}\right\}}{1-\;\mathbf{P}\left\{\bigcup\limits_{k=n+1}^{\infty}\{S_k > 0\}\,|\,\bigcap\limits_{k=1}^{n}\{S_k \le 0\}\right\}}, \tag{4.2}$$

because $\mathbf{P}\{\tau_+ = \infty\} = \eta$ from Siegmund (1985). Consider the conditional probability that appear in the numerator and the denominator of (4.2). It is clear that for any $k \ge 1$ the indicators of the events $\bigcup\limits_{k=n+1}^{\infty}\{S_k > 0\}$ and $\bigcap\limits_{k=1}^{n}\{S_k \le 0\}$ are respectively non-decreasing and non-increasing functions of the initial i.i.d. variables $Z_1, Z_2, \ldots$. Therefore, for every fixed k the pair of these indicators is negatively quadrant dependent (see Lehmann (1966)). Hence, by the definition of negatively quadrant dependence,

$$\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\},\,\bigcap_{k=1}^{n}\{S_k \le 0\}\right\} \le \;\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\}\right\}\;\mathbf{P}\left\{\bigcap_{k=1}^{n}\{S_k \le 0\}\right\},$$

which leads to

$$\begin{aligned}&\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\}\,|\,\bigcap_{k=1}^{n}\{S_k \le 0\}\right\} \\ &\qquad\le\;\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > 0\}\right\} \le \sum_{k=n+1}^{\infty}\;\mathbf{P}\{S_k > 0\}.\end{aligned} \tag{4.3}$$

By Chernoff theorem (see Chernoff (1952)), $\mathbf{P}\{S_k > 0\} \le \sigma^k$, which implies $\sum\limits_{k=n+1}^{\infty}\mathbf{P}\{S_k > 0\} \le \dfrac{\sigma^{n+1}}{1-\sigma}$ as $n \to \infty$, and (4.1) follows from (4.2) and (4.3).

Estimation by a geometric series, however, becomes useless when it results in a bound for (4.3) greater than 1. This happens if $\sigma^{n+1}/(1-\sigma) > 1$, or $n > \log(1-\sigma)/\log\sigma - 1$.

□

Lemma 4.2 *For any positive c and n*

$$\mathbf{P}\{n < t_c < \infty\} \le \;\mathbf{P}\{t_c = \infty\}\frac{\sigma^{n+1}}{1-\sigma-\sigma^{n+1}} \tag{4.4}$$

if $n > \log(1-\sigma)/\log\sigma - 1$.

Proof: Similarly to the proof of Lemma 4.1,

$$\mathbf{P}\{n < t_c < \infty\} \;=\; \frac{\mathbf{P}\{t_c = \infty\}\,\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > c\} \mid \bigcap_{k=1}^{n}\{S_k \le c\}\right\}}{1 - \;\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > c\} \mid \bigcap_{k=1}^{n}\{S_k \le c\}\right\}},$$

and (4.4) follows from Chernoff theorem, because

$$\mathbf{P}\left\{\bigcup_{k=n+1}^{\infty}\{S_k > c\} \mid \bigcap_{k=1}^{n}\{S_k \le c\}\right\} \;\le\; \sum_{k=n+1}^{\infty}\mathbf{P}\{S_k > c\} \;\le\; \sum_{k=n+1}^{\infty}\sigma^k.$$

$\square$

One can obtain a better upper bound for $\mathbf{P}\{n < t_c < \infty\}$ since Chernoff theorem guarantees

$$\mathbf{P}^{1/k}\{S_k > c\} \;\le\; \inf_s \exp\left\{-s\frac{c}{k}\right\}\,\mathbf{E}e^{Z_1 s} \;\le\; \sigma.$$

In this chapter, however, we study asymptotic behavior of the confidence set characteristics as $n \to \infty$, and this improvement is negligible for the large values of n.

Lemma 4.1 follows from Lemma 4.2 if we put $c = 0$. It will be used for obtaining the convergence rate of error probabilities. For the application to the confidence sets R_c we need to consider large values of c. Then one can use (2.5) to approximate $\mathbf{P}\{t_c = \infty\}$ when $Z_1, Z_2, \ldots$ are likelihood ratios.

4.1.1 Negative Binomial process

More accurate results can be obtained when $Z_1 = 1$ with probability p, $0 < p < 1/2$, and $Z_1 = -1$ with probability $q = 1 - p$. Then S_k is called the Negative

Binomial process (cf. Gut (1988), Chapter II). In that case the exact asymptotics for the probabilities (4.1) and (4.4) is available.

Let $\lceil x \rceil$ denote the smallest integer which is not smaller than x.

Lemma 4.3 *Let S_k be a Negative Binomial process with parameter p, $0 < p < 1/2$, and the unit span. Then, for any positive integer-valued function $c = c(n)$ such that $n - c \to \infty$ as $n \to \infty$,*

$$\mathbf{P}\{n < t_c < \infty\} \sim \frac{2^{3/2}}{\sqrt{\pi}} \frac{q\,(c+1)}{(1-4pq)(2p)^c} \frac{(4pq)^{\lceil \frac{n+c}{2} \rceil}}{n^{3/2}}. \tag{4.5}$$

Proof: Let $N(l,m;x,y)$ denote the cardinality of the set $\mathcal{A}$ of all possible realizations of a segment of random walk S_j from $j = l$ to $j = m$ such that $S_l = x$ and $S_m = y$, and let $N_z(l,m;x,y)$ be the number of elements of $\mathcal{A}$ satisfying the condition

$S_{l+1} \le z, \ldots, S_{m-1} \le z$.

Then, for any positive integer k and c such that $k+c$ is odd, according to the reflection principle (cf. Feller (1966)), one has

$$\begin{aligned} N_c(0,k;0,c+1) &= N_c(0,k-1;0,c) \\ &= N(0,k-1;0,c) \; - \; N(0,k-1;0,c+2) \\ &= \binom{k-1}{\frac{k+c-1}{2}} - \binom{k-1}{\frac{k+c+1}{2}}. \end{aligned}$$

Also, it is clear that $N_c(0,k;0,c+1)$ vanishes if $k+c$ is even. Hence, with $\sum^*$ denoting the summation over the set of k where $k+c$ is odd,

$$\begin{aligned} \mathbf{P}\{n < t_c < \infty\} &= \sum_{k=n+1}^{\infty} \mathbf{P}\{t_c = k\} \\ &= \sum_{k=n+1}^{\infty} \mathbf{P}\{S_1 \le c, \cdots, S_{k-1} \le c,\; S_k = c+1\} \end{aligned}$$

$$= \sum_{k=n+1}^{\infty}{}^{*} N_c(0,k;0,c+1)p^{(k+c+1)/2}q^{(k-c-1)/2}$$

$$= \sum_{j=\lceil \frac{n+c}{2} \rceil}^{\infty} \left\{ \binom{2j-c}{j} - \binom{2j-c}{j+1} \right\} p^{j+1}q^{j-c}. \tag{4.6}$$

According to Stirling formula, as $j \to \infty$,

$$\binom{2j-c}{j} - \binom{2j-c}{j+1} = \frac{(2j-c)!}{(j+1)!(j-c)!}(c+1) \sim$$
$$\sim \frac{e}{\sqrt{\pi}} \frac{(2j-c)^{2j-c}}{(j+1)^{j+1}(j-c)^{j-c}} \sqrt{\frac{2j-c}{(j+1)(j-c)}}\,(c+1) \sim$$
$$\sim \frac{e\,(c+1)}{\sqrt{\pi}}\, 2^{2j-c} \left(\frac{j-c/2}{j+1}\right)^{j+1} \left(\frac{j-c/2}{j-c}\right)^{j-c} \frac{1}{(j-c/2)\sqrt{j+1}} \sim$$
$$\sim \frac{c+1}{\sqrt{\pi}} \frac{2^{2j-c}}{(j-c/2)\sqrt{j+1}}.$$

Then from (4.6) one has as $n \to \infty$

$$\mathbf{P}\{n < t_c < \infty\} \sim \frac{c+1}{\sqrt{\pi}} \frac{q}{(2p)^c} \sum_{j=\lceil \frac{n+c}{2} \rceil}^{\infty} \frac{(4pq)^j}{(j-c/2)\sqrt{j+1}}. \tag{4.7}$$

Denote for simplicity $r = 4pq$, $s = \left\lceil \frac{n+c}{2} \right\rceil - 1$. Then

$$\sum_{j=s}^{\infty} \frac{r^j}{(j-c/2)\sqrt{j+1}} = \frac{r^s}{s^{3/2}} \sum_{j=0}^{\infty} \frac{r^j}{(1 - \frac{c}{2(j+s)})\sqrt{1 + \frac{1}{j+s}}} = \frac{r^s}{s^{3/2}} \sum_{j=0}^{\infty} a_j(s)\, r^j.$$

Since $a_j(s) \le 2$ for all $s \ge c$, and $r < 1$, the limit of $\sum a_j(s)\, r^j$ as $s \to \infty$ can be found by the Lebesgue Dominated Convergence theorem. Then the sum in (4.7) is equivalent to $\dfrac{r^s}{s^{3/2}(1-r)}$, and substituting these results in (4.7), one obtains the desired asymptotics. $\square$

Setting $c = 0$ in (4.5) one has an immediate corollary for the tail probabilities of the first ascending ladder epoch for Negative Binomial process.

Corollary 4.4 *As* $n \to \infty$,

$$\mathbf{P}\{n < \tau_+ < \infty\} \;\sim\; \frac{2^{3/2}}{\sqrt{\pi}} \frac{q}{1-4pq} \frac{(4pq)^{\lceil \frac{n}{2} \rceil}}{n^{3/2}}. \tag{4.8}$$

Remark. Easy calculation shows that $\sigma = 2\sqrt{pq}$ for Bernoulli distributions. Therefore Lemma 4.3 and Corollary 4.4 agree with the general results (4.1) and (4.4). However, the presence of the term $n^{-3/2}$ in (4.5) and (4.8) implies that inequalities (4.1) and (4.4) can be improved, although not in the exponential order.

4.2 Probabilities of the correct decision

The goal of this section is to compare the probabilities of hitting the true value of the change point ν by maximum likelihood estimators, calculated on the base of a finite sample and on the base of an infinite sample. For this purpose we consider a sequence of change-point estimators $\hat{\nu}_n$, which are obtained by a traditional maximum likelihood procedure, based on a finite random sample $\{X_0, \cdots, X_n\}$. For our convenience, we enumerate the data from 0 to n. The results of this section do not depend on the way of enumeration. To get a more commonly used symmetric sample, one only has to subtract $[n/2]$ from all the indices.

The true value of the parameter ν and $n - \nu$ are assumed to increase to ∞ as $n \to \infty$. Let $\hat{\nu}_\infty$ be the maximum likelihood estimator of ν based on the infinite sample $\{\cdots, X_{-1}, X_0, X_1, \cdots\}$. In our notations,

$$\hat{\nu}_\infty = \arg \max_{-\infty < k < \infty} S_k = \arg\left(\max_{0 \le k < \infty} S_k^{(1)} \vee \max_{0 < k < \infty} S_k^{(2)} \right),$$

where $S_k^{(1)} = S_{\nu - k} - S_\nu$ and $S_k^{(2)} = S_{\nu + k} - S_\nu$ for $k \ge 0$. It is known from Rukhin (1994) that

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}\{\hat{\nu}_n = \nu\} \;&=\; \mathbf{P}\{\hat{\nu}_\infty = \nu\} \\ &=\; \mathbf{P}\{\tau_+^{(1)} = \infty\}\, \mathbf{P}\{\tau_+^{(2)} = \infty\} \;= \eta_1 \eta_2, \end{aligned} \tag{4.9}$$

where, as in the previous chapters, $\tau_+^{(j)} = \inf\{k > 0 : \ S_k^{(j)} > 0\}$ and $\eta_j = \exp(-\sum_1^\infty k^{-1}\,\mathbf{P}\{S_k^{(j)} > 0\})$ for $j = 1, 2$.

In this section we prove that the rate of convergence in (4.9), which may be also considered as the rate of convergence of error probabilities, is exponential in $\min\{\nu, n-\nu\}$. The exact asymptotics is derived for the case of Bernoulli distributions.

Suppose that the estimators $\hat\nu_n$ and $\hat\nu_\infty$ were obtained from the same data set $\{X_k\}_{-\infty}^\infty$, but $\hat\nu_n$ was based on the subsample $\{X_0, \cdots, X_n\}$ only. Since it is assumed that $0 \le \nu \le n$, it is more likely to hit the true value of ν by using $\hat\nu_n$ than by using $\hat\nu_\infty$. More precisely,

$$
\begin{aligned}
\mathbf{P}\{\hat\nu_n = \nu\} - \mathbf{P}\{\hat\nu_\infty = \nu\} &= \mathbf{P}\{S_\nu = \max_{0\le k\le n} S_k < \max_{-\infty<k<\infty} S_k\} \\
&= \mathbf{P}\{\tau_+^{(1)} > \nu,\ \tau_+^{(2)} > n-\nu,\ \min\{\tau_+^{(1)}, \tau_+^{(2)}\} < \infty\} \\
&= \mathbf{P}\left\{\begin{matrix} \nu < \tau_+^{(1)} < \infty \\ \tau_+^{(2)} = \infty \end{matrix}\right\} + \mathbf{P}\left\{\begin{matrix} \tau_+^{(1)} = \infty \\ n-\nu < \tau_+^{(2)} < \infty \end{matrix}\right\} + \mathbf{P}\left\{\begin{matrix} \nu < \tau_+^{(1)} < \infty \\ n-\nu < \tau_+^{(2)} < \infty \end{matrix}\right\} \\
&= \eta_2\,\mathbf{P}\{\nu < \tau_+^{(1)} < \infty\} + \eta_1\,\mathbf{P}\{n-\nu < \tau_+^{(2)} < \infty\} \\
&\qquad + \mathbf{P}\{\nu < \tau_+^{(1)} < \infty\}\,\mathbf{P}\{n-\nu < \tau_+^{(2)} < \infty\}.
\end{aligned} \tag{4.10}
$$

Here we used independence of $\tau_+^{(1)}$ and $\tau_+^{(2)}$ and the fact that $\mathbf{P}\{\tau_+^{(j)} = \infty\} = \eta_j$ for $j = 1, 2$. Formula (4.10) relates the difference of error probabilities to the tail probabilities of $\tau_+^{(1)}$ and $\tau_+^{(2)}$. In the next section, the rates of convergence of the coverage probability and the expected width of confidence estimators are also derived through these quantities. The asymptotic behavior of the probabilities of this kind was studied in Section 4.1. To use our results, we note that if a random walk is generated by likelihood ratios $Z_j = \log \frac{f}{g}(X_j)$, then

$$
\sigma_1 = \inf_{0\le s\le 1} \mathbf{E}\exp\{S_1^{(1)} s\} = \inf_s \mathbf{E}_F \left(\frac{g}{f}(X)\right)^S = \inf_s \int f^{1-s} g^s
$$

$$= \quad \inf_s \int f^s g^{1-s} \;=\; \sigma_2 \;= \sigma \qquad (4.11)$$

is the Chernoff entropy. Also note that one can use (2.9) to evaluate $\eta = \eta_1 = \eta_2$ for of Bernoulli distributions,

$$\eta = 1 - \frac{p}{q}\,.$$

After certain simplification, from (4.1), (4.8) and (4.10) one obtains the following

Proposition 4.5 *As $\nu \to \infty$ and $n - \nu \to \infty$,*

$$\mathbf{P}\{\hat{\nu}_n = \nu\} - \mathbf{P}\{\hat{\nu}_\infty = \nu\} \le \eta_1\eta_2\frac{\sigma}{1-\sigma}(\sigma^\nu + \sigma^{n-\nu})(1 + o(1)),$$

where σ is the Chernoff entropy given by (4.11).

If F and G are Bernoulli distributions with parameters p, $0 < p < 1/2$ or $1/2 < p < 1$, and $q = 1 - p$ respectively, then

$$\mathbf{P}\{\hat{\nu}_n = \nu\} - \mathbf{P}\{\hat{\nu}_\infty = \nu\} \sim \frac{2^{3/2}}{\sqrt{\pi}\,|p-q|}\left\{\frac{(4pq)^{\lceil \frac{\nu}{2} \rceil}}{\nu^{3/2}} + \frac{(4pq)^{\lceil \frac{n-\nu}{2} \rceil}}{(n-\nu)^{3/2}}\right\}.$$

4.3 Coverage probabilities and expected width

In this section we consider the problem of a confidence estimation in the similar aspect. Suppose that $R_c(n)$ and $R_c(\infty)$ are the traditional confidence regions which were obtained from the sample of a size n and the infinite sample respectively. Namely, if an infinite sample $\{\ldots, X_{-1}, X_0, X_1, \ldots\}$ was observed, let

$$\begin{aligned} R_c(n) &= \{k \in [0; n] \,:\, S_{\hat{\nu}_n} - S_k \le c\}, \\ R_c(\infty) &= \{k \in (-\infty; \infty) \,:\, S_{\hat{\nu}_\infty} - S_k \le c\} \end{aligned}$$

for some positive constant c, independent on n.

It follows from Chapter 2 that $\mathbf{P}_\nu\{\nu \in R_c(n)\} \to \mathbf{P}_\nu\{\nu \in R_c(\infty)\}$ and $\mathbf{E}_\nu|R_c(n)| \to \mathbf{E}_\nu|R_c(\infty)|$ as $n \to \infty$. Our next goal is to study the rate of this convergence.

For all n consider two sets:

$$\begin{aligned} U_n &= \{k \in [0;n] : S_{\hat\nu_n} - c \le S_k < S_{\hat\nu_\infty} - c\}, \\ V_n &= \{k \notin [0;n] : S_k \ge S_{\hat\nu_\infty} - c\}. \end{aligned}$$

Note that $U_n \subset R_c(n)$ and $V_n \cap R_c(n) = \emptyset$. Also,

$$R_c(\infty) = (R_c(\infty) \cap [0;n]) \cup (R_c(\infty) \setminus [0;n]) = (R_c(n) \setminus U_n) \cup V_n. \tag{4.12}$$

Since $0 \le \nu \le n$, one has $\mathbf{P}\{\nu \in V_n\} = 0$ for all n, and therefore

$$\begin{aligned} &\mathbf{P}_\nu\{\nu \in R_c(n)\} - \mathbf{P}_\nu\{\nu \in R_c(\infty)\} = \mathbf{P}_\nu\{\nu \in U_n\} \\ &= \mathbf{P}\{\nu < t_c^{(1)} < \infty\}\,\mathbf{P}\{t_c^{(2)} > n-\nu\} + \mathbf{P}\{t_c^{(1)} > \nu\}\,\mathbf{P}\{n-\nu < t_c^{(2)} < \infty\} - \\ &\quad - \mathbf{P}\{\nu < t_c^{(1)} < \infty\}\,\mathbf{P}\{n-\nu < t_c^{(2)} < \infty\} \\ &= \mathbf{P}\{t_c^{(1)} > \nu\}\,\mathbf{P}\{t_c^{(2)} > n-\nu\} - \mathbf{P}\{t_c^{(1)} = \infty\}\,\mathbf{P}\{t_c^{(2)} = \infty\}, \end{aligned} \tag{4.13}$$

because

$$\begin{aligned} \{\nu \in U_n\} = &\left(\{\max_{[\nu+1;\infty]} S_k^{(1)} > c\} \cup \{\max_{[n-\nu+1;\infty]} S_k^{(2)} > c\}\right) \bigcap \\ &\bigcap \left(\{\max_{[0;\nu]} S_k^{(1)} \le c\} \cap \{\max_{[0;n-\nu]} S_k^{(2)} \le c\}\right). \end{aligned}$$

The asymptotic behavior of the probabilities of the form $\mathbf{P}\{t_c^{(j)} > t\}$ as $t \to \infty$ follows from Lemma 4.2 for the general case and from Lemma 4.3 for Bernoulli distributions. Applying these results to (4.13), one has for $\nu \to \infty$ and $n-\nu \to \infty$

$$\begin{aligned} &\mathbf{P}_\nu\{\nu \in R_c(n)\} - \mathbf{P}_\nu\{\nu \in R_c(\infty)\} \\ &\sim \mathbf{P}\{t_c^{(1)} = \infty\}\,\mathbf{P}\{n-\nu < t_c^{(2)} < \infty\} + \mathbf{P}\{t_c^{(2)} = \infty\}\,\mathbf{P}\{\nu < t_c^{(1)} < \infty\} \end{aligned} \tag{4.14}$$

$$\leq \quad \mathbf{P}\{t_c^{(1)} = \infty\}\, \mathbf{P}\{t_c^{(2)} = \infty\} \left(\frac{\sigma^{n-\nu+1}}{1-\sigma-\sigma^{n-\nu+1}} + \frac{\sigma^{\nu+1}}{1-\sigma-\sigma^{\nu+1}} \right). \qquad (4.15)$$

For large values of c one can use (2.5) to approximate $\mathbf{P}\{t_c^{(j)} = \infty\}$, $j = 1, 2$. According to this formula, from (4.15),

$$\begin{aligned} &\mathbf{P}_\nu\{\nu \in R_c(n)\} - \mathbf{P}_\nu\{\nu \in R_c(\infty)\} \\ &\quad \leq \left(1 - \frac{e^{-c}\eta_1\eta_2'}{\rho_1}\right)\left(1 - \frac{e^{-c}\eta_1'\eta_2}{\rho_2}\right) \frac{\sigma^{\nu+1}+\sigma^{n-\nu+1}}{1-\sigma}(1+o(1)) \end{aligned}$$

for sufficiently large values of c, as $n \to \infty$.

In the case of Bernoulli distributions with parameters p and $q = 1-p$, $S_k^{(1)}$ and $S_k^{(2)}$ are identically distributed Negative Binomial processes with a span $\zeta = |\log(p/q)|$. Hence, for any c that is an integer multiple of ζ, one has to divide c by ζ in the right-hand side of (4.5) to get a correct asymptotics of $\mathbf{P}\{\nu < t_c^{(1)} < \infty\}$ and $\mathbf{P}\{n-\nu < t_c^{(2)} < \infty\}$. Also, using (2.8) one has

$$\mathbf{P}\{t_c = \infty\} = \mathbf{P}\{\max_{k\geq 0} S_k \leq c\} = \mathbf{P}\{\max_{k\geq 0} S_k < c+\zeta\} = 1 - \exp\{-(c+\zeta)\}.$$

Finally, using this to find the asymptotics of (4.14) for Negative Binomial process, one has

$$\begin{aligned} &\mathbf{P}_\nu\{\nu \in R_c(n)\} - \mathbf{P}_\nu\{\nu \in R_c(\infty)\} \\ &\quad \sim \frac{2^{3/2}}{\sqrt{\pi}} \frac{q(1-e^{-(c+\zeta)})}{1-4pq} \frac{c/\zeta+1}{(2p)^{c/\zeta}} \left(\frac{(4pq)^{\lceil \frac{\nu+c/\zeta}{2} \rceil}}{\nu^{3/2}} + \frac{(4pq)^{\lceil \frac{n-\nu+c/\zeta}{2} \rceil}}{(n-\nu)^{3/2}} \right) \end{aligned}$$

as $n \to \infty$.

Representation (4.12) can also be used to study the rate of convergence of the expected size $\mathbf{E}_\nu|R_c(n)|$. Indeed, since $|U_n| = \sum_k I\{k \in U_n\}$ and $|V_n| = \sum_k I\{k \in V_n\}$, from (4.12),

$$\begin{aligned} &\mathbf{E}_\nu|R_c(n)| - \mathbf{E}_\nu|R_c(\infty)| = \mathbf{E}_\nu|U_n| - \mathbf{E}_\nu|V_n| = \\ &= \sum_{k=0}^{n} \mathbf{P}\{k \in U_n\} - \sum_{k<0} \mathbf{P}\{k \in V_n\} - \sum_{k>n} \mathbf{P}\{k \in V_n\}. \end{aligned}$$

Considering three terms in (4.16) separately, one has for any $0 \le k \le n$

$$\begin{aligned} \mathbf{P}\{k \in U_n\} &\le \mathbf{P}\{\hat{\nu}_n \ne \hat{\nu}_\infty\} \\ &\le \sum_{j<0} \mathbf{P}\{j = \hat{\nu}_\infty\} + \sum_{j>n} \mathbf{P}\{j = \hat{\nu}_\infty\} \\ &\le \sum_{j=\nu+1}^{\infty} \mathbf{P}\{S_j^{(1)} > 0\} + \sum_{j=n-\nu+1}^{\infty} \mathbf{P}\{S_j^{(2)} > 0\} \\ &\le \sum_{\nu+1}^{\infty} \sigma^j + \sum_{n-\nu+1}^{\infty} \sigma^j = \frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma}, \end{aligned}$$

from where

$$\mathbf{E}_\nu |U_n| \le \sum_{k=0}^{n} \frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma} = (n+1)\frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma}.$$

The other two terms in (4.16) admit even smaller asymptotic bounds, because

$$\mathbf{P}\{k \in V_n\} = \mathbf{P}\{S_{\nu-k}^{(1)} \ge S_{\hat{\nu}_\infty} - c\} \le \mathbf{P}\{S_{\nu-k}^{(1)} \ge 0\} \le \sigma^{\nu-k},$$

if $k < 0$, and similarly, $\mathbf{P}\{k \in V_n\} \le \sigma^{k-\nu}$ for $k > n$. Hence,

$$0 \le \mathbf{E}_\nu |V_n| = \sum_{k<0} \mathbf{P}\{k \in V_n\} + \sum_{k>n} \mathbf{P}\{k \in V_n\} \le \frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma}.$$

Therefore, for sufficiently large n

$$|\, \mathbf{E}_\nu |R_c(n)| - \mathbf{E}_\nu |R_c(\infty)|| = |\, \mathbf{E}|U_n| - \mathbf{E}|V_n|| \le (n+1)\frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma}.$$

We summarize the results of this section in the following theorem.

Theorem 4.6 *If* $\min\{\nu, n-\nu\} \to \infty$ *as* $n \to \infty$ *, then for sufficiently large values of* c

$$\begin{aligned} &\mathbf{P}_\nu\{\nu \in R_c(n)\} - \mathbf{P}_\nu\{\nu \in R_c(\infty)\} \\ &\quad \le \left(1 - \frac{e^{-c}\eta_1\eta_2'}{\rho_1}\right)\left(1 - \frac{e^{-c}\eta_1'\eta_2}{\rho_2}\right)\frac{\sigma^{\nu+1} + \sigma^{n-\nu+1}}{1-\sigma}(1+o(1)) \end{aligned}$$

as $n \to \infty$, *and*

$$|\ \mathbf{E}_\nu|R_c(n)| - \ \mathbf{E}_\nu|R_c(\infty)||\ \leq\ (n+1)\frac{\sigma^{\nu+1}+\sigma^{n-\nu+1}}{1-\sigma}.$$

For the case of Bernoulli distributions with parameters $0 < p < 1/2$ *and* $q = 1 - p$

$$\mathbf{P}_\nu\{\nu \in R_c(n)\} - \ \mathbf{P}_\nu\{\nu \in R_c(\infty)\}$$
$$\sim\ \frac{2^{3/2}}{\sqrt{\pi}}\,\frac{(q - p\,e^{-c})(c+1)}{(1-4pq)(2p)^c}\left(\frac{(4pq)^{\lceil\frac{\nu+c}{2}\rceil}}{\nu^{3/2}} + \frac{(4pq)^{\lceil\frac{n-\nu+c}{2}\rceil}}{(n-\nu)^{3/2}}\right)$$

for any c *as* $n \to \infty$.

It remains to notice that Theorem 3.4 guarantees exponential in a power of n rate of convergence of the coverage probability and of the expected width of the confidence set T_c to those of the set R_c. Hence, in the situation discussed in Chapter 3, when the distribution before and after the change point belong to an exponential family with an unknown nuisance parameter, one can use the obtained asymptotic formulae for finite data sets with the approximation error tending to zero exponentially fast.

Chapter 5

Application to the data of measurement of mass standards

As a practical example of confidence estimation of the change-point parameter, we present the analysis of a sequence of precision measurements of mass standards made at the National Institute of Standards and Technology (NIST) between 1975 and 1988.

The calibration process of mass standards at the NIST consists of inter-comparison measurements of two NIST 1 kilogram standards and two client's weights, 1 kg and a set of 500, 300 and 200 g denominations. All six differences between these four weights, say, $w_1, w_2, \ldots, w_6$, are recorded, and the further analysis is based on them. Any significant change in the sequence of data invalidates the calibration process.

Assuming a linear model of the form

$$\begin{cases} w_1 &= \mu_1 + \epsilon_1 \\ w_2 &= \mu_2 + \epsilon_2 \\ w_3 &= \mu_3 + \epsilon_3 \\ w_4 &= \mu_2 - \mu_1 + \epsilon_4 \\ w_5 &= \mu_3 - \mu_1 + \epsilon_5 \\ w_6 &= \mu_3 - \mu_2 + \epsilon_6 \end{cases}$$

with ϵ_j being independent and identically distributed normal random variables with mean 0 and an unknown variance σ_ϵ^2, the difference between the two NIST kilograms μ_1 is estimated by the method of least squares. This estimate, X, is used as a check standard value for the calibration process.

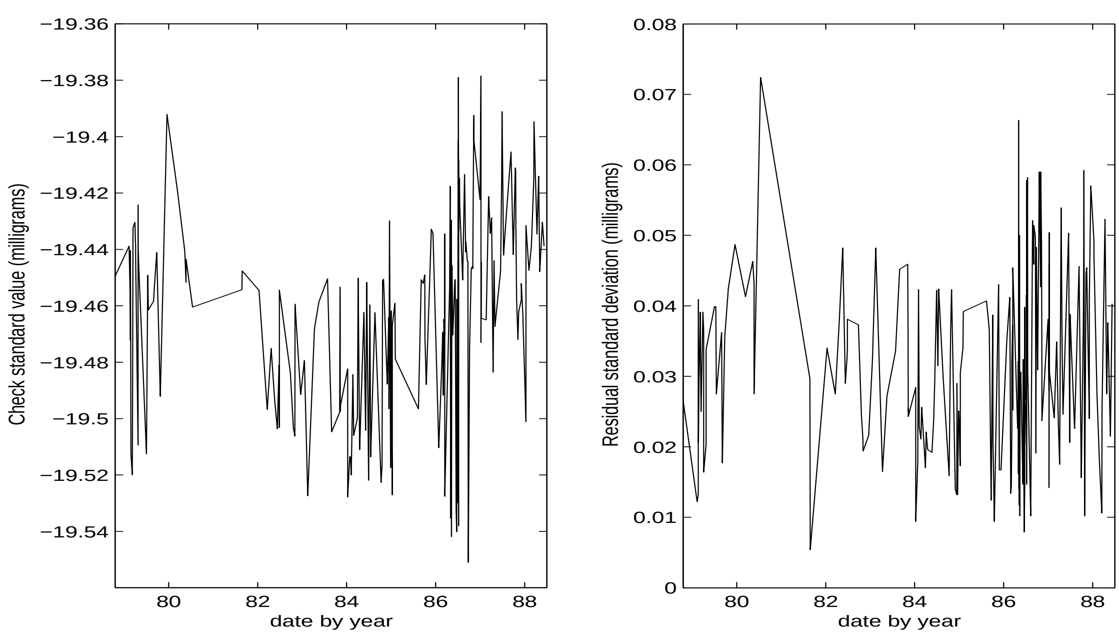


Figure 5.1: *217 check standard values X and residual standard deviations Y estimated from a comparison design at NIST between 1975 and 1988.*

Also the residual standard deviation

$$Y = \sqrt{\frac{1}{3}\sum_{j=1}^{6}(w_j - \hat{w}_j)^2} \tag{5.1}$$

with $\hat{w}_j$ being the least squares estimates of $\mathbf{E}w_j$ for $j = 1, \ldots, 6$ is calculated from the same design. That is, X is the statistic for monitoring the process mean, and Y is the statistic for monitoring the process precision. The values of X and Y, in milligrams, observed $n = 217$ times between 1975 and 1988, are depicted in Figure 5.1.

The Shewhart $\bar{X}$-charts, normally used for this procedure, appeared to be ineffective in detecting and estimating the change point in the process. The analysis presented below indicates a significant change in the distribution of both X and Y between the observations 145 and 160, that is, in the summer of 1986. We study the two sets of data separately because under the assumptions of the model X and Y are independent, and this is supported by a small value of the correlation coefficient, $r_{XY} = 0.0924$.

Analysis of the check standard value. Following the procedure described in Chapter 3, let $\hat{\theta}_1(k)$ and $\hat{\theta}_2(k)$ be the maximum likelihood estimators for a two-dimensional parameter $\theta = (\mu, \sigma^2)$ of the normal distribution of X for each hypothetical value k of the change point ν. That is, if for any k and m

$$\bar{X}_{k:m} = \frac{1}{m-k+1}\sum_{i=k}^{m} X_i \quad \text{and} \quad S^2_{k:m} = \frac{1}{m-k+1}\sum_{i=k}^{m}(X_i - \bar{X}_{k:m})^2,$$

then we have

$$\hat{\theta}_1(k) = (\bar{X}_{1:k},\, S^2_{1:k}) \quad \text{and} \quad \hat{\theta}_2(k) = (\bar{X}_{k+1:n},\, S^2_{k+1:n}).$$

(We note here that although our analysis in chapter 3 concerns only one-dimensional parameter exponential families, the generalization to higher dimensions is straightforward. One only has to write multivariate Taylor expansions in (3.9) and (3.17).)

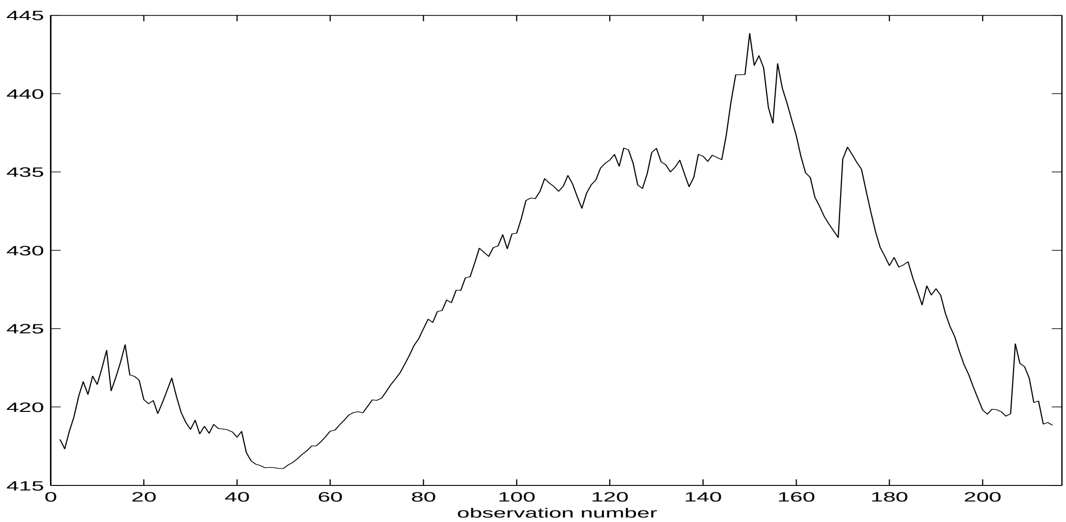
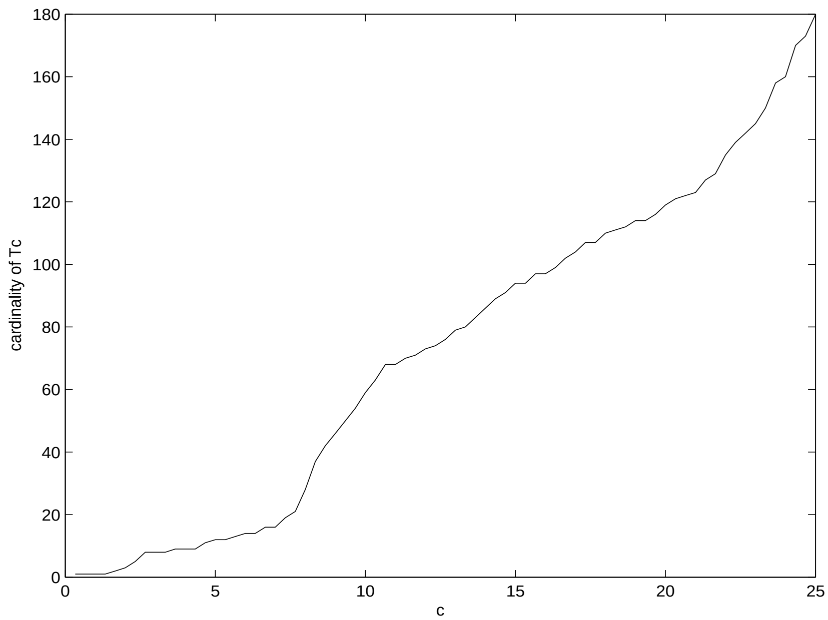


Figure 5.2: *Log-likelihood process with estimated nuisance parameters for the check standard values X.*

Substituting the estimates of the nuisance parameters into the log-likelihood function of the distribution of X, one obtains the following expression for the log-likelihood process Λ_k (see Figure 5.2)

$$\Lambda_k = -\frac{n}{2}\log(2\pi e) - k\log S_{1:k} - (n-k)\log S_{k+1:n}.$$

At the observation number $k = 150$ the process Λ_k attains its extremum, so that $\hat{\nu} = 150$ is the overall maximum likelihood estimator of the change-point parameter. For different values of c the confidence set T_c includes all the time points k where $\Lambda_k > \Lambda_{150} - c \approx 443.84 - c$. Figure 5.3 shows the dependence of the number of elements in T_c on the size of the threshold c.



c	$\lvert T_c \rvert$	c	$\lvert T_c \rvert$
1	1	10	59
2	3	11	68
3	8	12	73
4	9	13	79
5	12	14	86
6	14	15	94
7	16	16	97
8	28	17	104
9	46	18	110

Figure 5.3: *Number of elements in the confidence set T_c for check standard value as a function of c.*

Analysis of the residual standard deviation. Under the assumption of normality and in view of (5.1), $3Y^2/\sigma_\epsilon^2$ follows a chi-square distribution with 3 degrees of freedom. Hence, Y has a chi density with a scale parameter $\sigma_\epsilon/\sqrt{3}$

$$f(y|\sigma_\epsilon) \; = \; 6\sqrt{\frac{3}{2\pi}}\,\frac{y^2}{\sigma_\epsilon^3}\,\exp\left\{-\frac{3y^2}{2\sigma_\epsilon^2}\right\}.$$

Then, with maximum likelihood estimators of σ_ϵ, $\hat{\sigma_\epsilon}^{(1)}(k) = \sqrt{\sum_1^k Y_j^2/k}$ and $\hat{\sigma_\epsilon}^{(2)}(k) = \sqrt{\sum_{k+1}^n Y_j^2/(n-k)}$ one has the following log-likelihood process (see Figure 5.4)

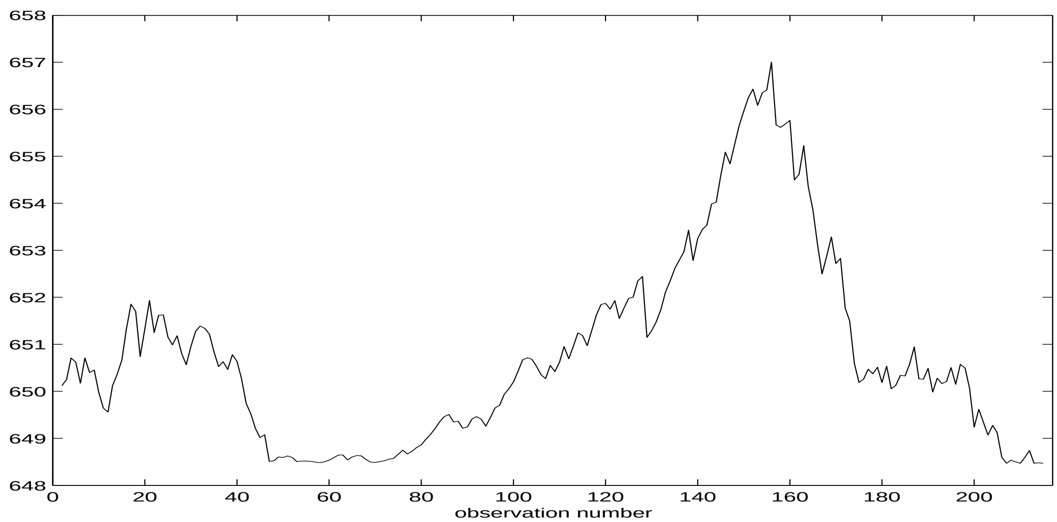
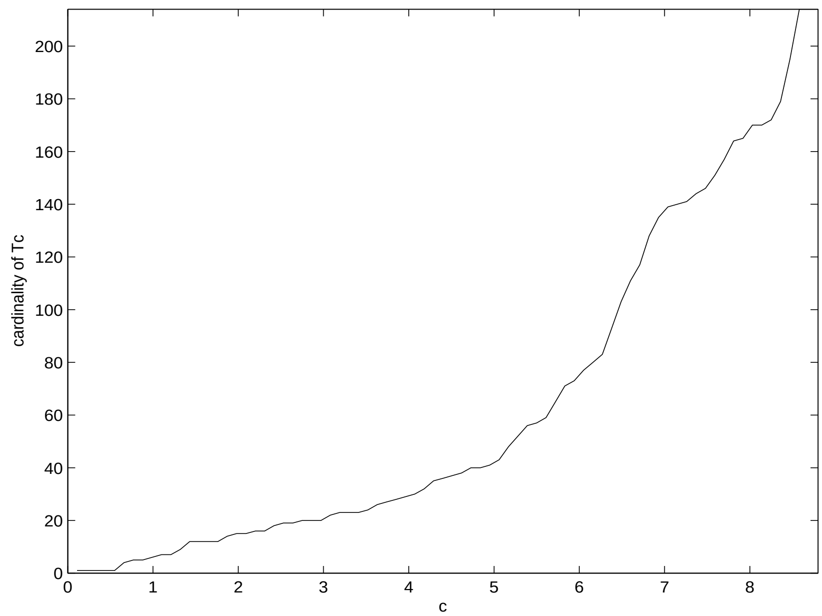


Figure 5.4: *Log-likelihood process with estimated nuisance parameters for the residual standard deviations Y.*

$$\Lambda_k = n\left\{\log\left(6\sqrt{\frac{3}{2\pi}}\right) - \frac{3}{2}\right\} + 2\sum_{j=1}^{n}\log Y_j - \frac{3}{2}\left\{k\log\frac{\sum_1^k Y_j^2}{k} + (n-k)\log\frac{\sum_{k+1}^n Y_j^2}{n-k}\right\}$$

Based on the data set Y, the overall maximum likelihood estimator of the change point is $\hat{\nu} = 156$. Looking at the graph of Λ_k for Y, one can notice another possible change point around 20th observation, however, the lack of data before this time point leaves this doubtful. Finally, we construct the confidence sets T_c for estimating ν from the data set Y by taking all the points k for which $\Lambda_k > \Lambda_{\hat{\nu}} - c \approx 657.0 - c$.



c	$\lvert T_c\rvert$	c	$\lvert T_c\rvert$
0.4	1	4.0	29
0.8	5	4.4	36
1.2	7	4.8	40
1.6	12	5.2	48
2.0	15	5.6	59
2.4	17	6.0	73
2.8	20	6.4	94
3.2	23	6.8	125
3.6	26	7.2	140

Figure 5.5: *Number of elements in the confidence set T_c for residual standard deviations as a function of c.*

Numerical procedure. Theorem 2.1 gives a way to construct approximately $(1-\alpha)100\%$ confidence sets for the change-point parameter for different confidence levels α.

First one has to evaluate the quantities ρ_j, η_j and η_j' for $j = 1, 2$. Clearly, $\eta_j = \eta_j'$ for continuous distributions. Also, when the distributions F and G belong to an exponential family with nuisance parameters θ_1 and θ_2 respectively, ρ_j and η_j are functions of θ_j. Hence, one can use overall maximum likelihood estimators of θ_j,

$$(\hat{\theta}_1, \hat{\theta}_2) \;=\; \arg\max_{\theta_1, \theta_2} \prod_{j=1}^{\hat{\nu}} f(X_j|\theta_1) \prod_{j=\hat{\nu}+1}^{n} f(X_j|\theta_2),$$

to evaluate the maximum likelihood estimators of ρ_j and η_j. In application to the NIST data, based on the data set X (check standard value), coming from the normal distribution, the following estimators are obtained,

$$\hat{\rho}_1 = \hat{\rho}_2 = 0.7229, \quad \hat{\eta}_1 = \hat{\eta}_2 = 0.6008.$$

For the random sample Y (residual standard deviation) from chi distribution with three degrees of freedom,

$$\hat{\rho}_1 = 0.1607,\ \ \hat{\rho}_2 = 0.2245,\ \ \hat{\eta}_1 = 0.4411,\ \ \hat{\eta}_2 = 0.2910.$$

Next, for different confidence levels α one can use (2.6) to evaluate the corresponding values of the threshold c and to construct the confidence regions. Following the results of Chapter 3, the approximation error due to the usage of maximum likelihood estimators is exponentially small for large sample sizes. The 90%, 95% and 99% confidence sets for the change point in the NIST data, based on the data sets X and Y, are given below.

<u>Based on check standard values X</u>

90%	$T_{2.3012}$	$= \{150,\ 151,\ 152,\ 153,\ 156\}$	$-$ 5 elements
95%	$T_{2.9943}$	$= \{147,\ 148,\ \ldots\ 153,\ 156\}$	$-$ 8 elements
99%	$T_{4.6037}$	$= \{146,\ 147,\ \ldots\ 153,\ 156,\ 157,\ 158\}$	$-$ 11 elements

<u>Based on residual standard deviations Y</u>

90%	$T_{2.6178}$	$= \{145,\ 146,\ \ldots\ 163\}$	$-$ 19 elements
95%	$T_{3.3109}$	$= \{143,\ 144,\ \ldots\ 165\}$	$-$ 23 elements
99%	$T_{4.9204}$	$= \{127,\ 128,\ 133,\ 134,\ \ldots\ 171\}$	$-$ 41 elements

Bibliography

[1] R. R. Bahadur. *Some Limit Theorems in Statistics.* Regional Conferences Series in Applied Mathematics. SIAM, Philadelphia, PA, 1971.

[2] D. Barry and J. Hartigan. Change points in 0-1 sequences, with an application to predicting divisional winners in major league baseball. *Journal of Applied Statistical Science*, 1:323–336, 1993.

[3] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric Methods in Change-Point Problems.* Kluwer Academic Publishers, The Netherlands, 1993.

[4] L. D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* Institute of Mathematical Statistics, Lecture Notes–Monograph Series, Volume 9, Hayward, CA, 1986.

[5] E. Carlstein. Nonparametric estimation of a change-point. *Ann. Statist.*, 16:188–197, 1988.

[6] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.

[7] H. Chernoff and S. Zacks. Estimating the current mean of a normal distribution which is subject to changes in time. *Ann.Math.Statist.*, 35:999–1028, 1964.

[8] G. W. Cobb. The problem of the Nile: conditional solution to a change-point problem. *Biometrika*, 62:243–251, 1978.

[9] B. S. Darkhovski. A non-parametric method for a posterior detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.*, 21:178–183, 1976.

[10] W. Feller. *An introduction to probability theory and its applications.* Wiley, New York, 1966.

[11] A. Gut. *Stopped random walks. Limit theorems and applications.* Springer-Verlag, New York, 1988.

[12] I. Hu and A. L. Rukhin. A lower bound for error probability in change-point estimation. *Statistica Sinica*, 5:319–331, 1995.

[13] B. James K. L. James and D. Siegmund. Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution. *Statist. Sinica*, 2:69–90, 1992.

[14] R. A. Khan. Wald's approximations to the average run length in cusum procedures. *J.Statist. Plann.*, 2:63–77, 1978.

[15] R. A. Khan. A sequential detection procedure and the related cusum procedure. *Sankhya*, 40:146–162, 1979a.

[16] R. A. Khan. Some first passage problems related to cusum procedures. *Stochastic processes and their applications*, 9:207–215, 1979b.

[17] E. L. Lehmann. Some concepts of dependence. *Ann.Math.Statist.*, 37:1137–1153, 1966.

[18] V. I. Lotov. On random walks in a strip. *Theory Probab. Appl.*, 1:165–170, 1991.

[19] I. B. MacNeill and Y. Mao. Change-point analysis for mortality and morbidity rate. *Journal of Applied Statistical Science*, 1:359–377, 1993.

[20] E. S. Page. Continuous inspection schemes. *Biomterika*, 41:100–115, 1954.

[21] E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biomterika*, 42:523–527, 1955.

[22] E. S. Page. On problems in which a change in a parameter occurs at an unknown point. *Biomterika*, 44:248–252, 1957.

[23] A. L. Rukhin. Asymptotic minimaxity in the change-point problem. In *D. Siegmund, E. Carlstein, and H. Mueller, editors, Proceedings on Change-Point Problems*, Institute of Mathematical Statistics, Hayward, CA, 1994.

[24] W. A. Shewhart. *The economic control of the quality of a manufactured product.* Van Nostrand, 1931.

[25] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals.* Springer-Verlag, New York, 1985.

[26] D. Siegmund. Confidence sets in change-point problems. *International Statistcal Review*, 56:31–48, 1988.

[27] A. F. M. Smith. A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62:407–416, 1975.

[28] K. J. Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73:91–104, 1986.