A NONPARAMETRIC CHANGE-POINT MODEL FOR PHASE II ANALYSIS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

QIQI DENG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DOUGLAS M. HAWKINS
ADVISOR

January 2009

Acknowledgements

I would like to express my deep and sincere gratitude to many people who made this thesis possible.

It is difficult to overstate my appreciation to my advisor Professor Douglas Hawkins. I am so grateful I had the opportunity to meet and start working with Doug in the summer three years ago. Since then, he has given me innumerous support and inspirations in my research. His perpetual enthusiasm and intellectual insights have motivated all his advisees, and I am lucky to be one of them. I owe him lots of gratitude for teaching me to look at the problem with an integral view, for giving me confidence and helping me out when I was frustrated, and for supporting whatever decision I made in my career. He may not even realize how much I have learned from him. Actually for me, he is the role model of a successful professor in the field.

I am also indebted to my committee members. I am delighted to work with Professor Snigdhansu Chatterjee as his research assistant for one year. He invited me to a new research field with many interesting topics. I remember all the challenges we had in that short time, as well as his assistance and guidance. My sincere thanks go to Professor Peihua Qiu, who made critical comments on my proposal and provide me valuable suggestions. I also want to thank Professor Wei Pan for giving me two semesters' wonderful courses and accepting to serve on my committee. It was such a pleasant and inspiring experience to debate with the whole committee.

I wish to extend my thanks to:

Professor Gary Oehlert, Glen Meeden, Yuhong Yang, Tiefeng Jiang and all other professors who give me encouragement and important guidance during my Ph.D. studies.

Dr. Dan Sargent and Dr. Wei Zhang for hiring me as intern at Mayo Clinic and Boehringer Ingelheim respectively. The associated experiences broadened my perspective and triggered my interest on the practical aspects in health care and pharmaceutical industry.

My dear schoolmates. You gave me inspiring discussions of the courses and research, as well as those sunny happy days in my memory.

Special thanks be to my boyfriend Yiqun Mou. I could not get through all the difficult times without your support and understanding. Thanks to my best friends Huizi Zhao, Lu Deng, Ziyi Sun, Linxi Li, Qiang Li, Yixiao Sheng, Fanhuan Zhou, Qin Yan and Yinan Liu, for all the emotional support and caring you provided. Your friendship is a gift I will cherish forever.

Lastly, and most importantly, I wish to thank my parents, Shilun Deng and Chaosu Yu, as well as my grandmother Quanfang Xie. Thank you for loving me and seeing me as "precious". Your unconditional love is the best thing I have ever got in my life. To them I dedicate this thesis.

Abstract

In statistical process control, the control chart based on the change-point model does not require prior knowledge about the parameters, making it an attractive technique. So far, change-point control charts are only developed under a normal assumption. But when the underlying distribution is not normal or unclear, this may not be appropriate. In this thesis, we propose a nonparametric change-point model based on the Mann-Whitney statistic for ongoing Phase II analysis, which has essentially the same computational complexity as the parametric. The simulated out-of-control ARL shows that this nonparametric model outperforms the parametric for small to moderate shifts, although loses for large shifts, even for data from normal distribution. And we show that modifying the parametric procedure to prohibit very short segments largely equalizes the performance of the parametric and nonparametric methods. Finally, an asymptotic limit of the control limit in parametric change-point model is proposed.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In statistical process control (SPC), one major concern is whether there has been a change of the distribution from the target in the process. To answer this question, we have to distinguish the variation due to real change of distribution (assignable causes) from that due to random error (chance causes). Many kinds of control charts serve this purpose. When there is no change and the process is in control (IC), the probability of a false signal from the chart should be controlled at a low and known false alarm rate. And when there is some change and the process is out of control (OOC), a good chart should detect it as soon as possible. The performance of a chart is usually measured by average run length (ARL) – the expected number of samples or sub-groups to be collected before the first signal.

Control charts often assume that data come from some parametric distribution, most commonly the normal distribution. When the underlying process is unknown or known not to be normal, these charts may not be appropriate. Woodall and Montgomery (1999), and Woodall (2000) have provided some discussion on this issue. Along with other literature, they motivate the development and use of nonparametric control charts. Chakraborti et al (2001) gave a comprehensive review of nonparamet-

ric control charts. Amin, Reynolds, and Bakir (1995) developed a Shewhart-type control charts assuming the location parameter (mean or median) is known. Some other Shewhart charts, for example, Willemain and Runger (1996), Janacek and Meikle (1997) were set up on some attribute of the in-control distribution, like the mean, median, percentile or inter-quartile range estimated from a reference sample. Bakir and Reynolds (1979) considered a CUSUM chart based on the Wilcoxon signed-rank (WSR) statistic. McDonald (1990) proposed a CUSUM based on "sequential ranks". These CUSUM methods incorporate the sequential nature of SPC and are thus effective in detecting small persistent changes. However, they require the knowledge of both the in-control and/or out-of control location parameters to set up the reference value. A EWMA-type chart was provided by Amin and Search (1991), who describe a GSR-EWMA chart based on a grouped signed-rank statistic (GSR). This chart too requires the in-control parameter to set up the starting value. Hackl and Ledolter (1991) used "standardized ranks" of the observations relative to an in-control distribution. When the in-control distribution is unknown, they suggest using the ranking on reference data to estimate it, but indicate that this estimation would lead to a substantially larger in-control ARL than nominal.

All these methods require the prior knowledge of the in-control parameter, or at least a large reference sample to get an adequate estimation for the parameter. Jones et al. (2001, 2004), Jensen et al. (2006) discussed the effect of parameter estimation on the performance of control chart, and pointed out that with a moderately-sized reference sample, control charts based on the estimated parameters will behave quite differently from desired. Specifically, estimation usually substantially increases the in-control ARL from the nominal value, and also reduces the sensitivity to change when the process goes out of control.

Recently, Hawkins, Qiu and Kang (2003) (shorthand HQK in this thesis) proposed a change-point control chart for normally distributed data. Instead of comparing

the subsequent observations with a known or estimated target value, their method treats the reference samples as part of the ongoing data stream and examines the consistency through the whole process. Lai (2001) listed a variety of change-point models for sequential detection, in which at least some in-control parameters were assumed to be known in advance. The HQK formulation however differs in not requiring advance knowledge about any parameter. The HQK method not only avoids the tenuous "known-parameter" assumption and thus always maintains the desired in-control ARL, but also seamlessly combines the data collection in Phase I and ongoing analysis in Phase II study.

Zhou et al (2007) developed a nonparametric change-point model motivated by HQK, but, deterred by issues of computational complexity, these authors did not follow the same methodology of HQK, but instead went to an exponentially weighted moving average (EWMA) method. The EWMA involves the additional element of requiring a tuning constant λ, the choice of which has a considerable effect on the performance, and negates some of the simplicity of the change-point formulation. Rather than sidestep the computational issue, in this thesis, we will tempt to overcome it, and develop a nonparametric parallel of HQK with modest computational needs.

Chapter 2

Change-point Model in Phase I analysis

2.1 Parametric Model

2.1.1 Existing Model

There is a substantial literature on the Phase I parametric change-point model. Suppose $X_1, X_2, \cdots, X_\tau, X_{\tau+1}, \cdots, X_n$ are independent, continuous random variables.

$$X_i \sim N(\mu, \sigma^2), \quad \text{for } i = 1, 2, \cdots, \tau; \tag{2.1.1}$$

$$X_i \sim N(\mu + \theta, \sigma^2), \quad \text{for } i = \tau + 1, \cdots, n. \tag{2.1.2}$$

To test whether the process is in control, the hypothesis is

$$H_o : \theta = 0, \qquad H_a : \theta \neq 0.$$

Again, the change point τ is an unknown parameter, and the sample size n is a fixed number. Without loss of generality, μ is set to be 0.

A method based on likelihood ratio test statistics is given in Hinkley (1970). In mathematical notation, let

$$\bar{X}_{0,k} = \sum_{i=1}^{k} X_i/k \tag{2.1.3}$$

and

$$\bar{X}_{k,n} = \sum_{i=k+1}^{n} X_i/(n-k) \tag{2.1.4}$$

be the mean of the first k and the last $n-k$ observations. When variance σ^2 is known, to test if there is a change, we use

$$Z_{\max,n} = \max_{1\le k\le n-1} |Z_{k,n}|, \tag{2.1.5}$$

where

$$Z_{k,n} = \sqrt{\frac{k(n-k)}{n}}\frac{\bar{X}_{0,k} - \bar{X}_{k,n}}{\sigma} \sim N(0,1), \tag{2.1.6}$$

and is basically the two sample Z test statistic between left and right sections of the observations calculated with respect to that change-point.

If $Z_{\max,n}$ is larger than the pre-specified control limit, it signals a change and the estimated change-point is

$$\hat{\tau}_Z = \text{argmax}_{1\le k\le n-1}|Z_{k,n}|.$$

When variance σ^2 is unknown, we use

$$\tilde{T}_{\max,n} = \max_{1\le k\le n-1} |\tilde{T}_{k,n}|, \tag{2.1.7}$$

where

$$\tilde{T}_{k,n} = \sqrt{\frac{k(n-k)}{n}}\frac{\bar{X}_{0,k} - \bar{X}_{k,n}}{\hat{\sigma}_{k,n}} \sim N(0,1), \tag{2.1.8}$$

and $\hat{\sigma}_{k,n}$ is the pooled within-segment standard deviation:

$$\hat{\sigma}_{k,n} = \frac{1}{n-2}[\sum_{i=1}^{k}(X_i - \bar{X}_{0,k})^2 + \sum_{i=k+1}^{n}(X_i - \bar{X}_{k,n})^2].$$

$\tilde{T}_{k,n}$ is basically the two sample T test statistic between left and right sections of the observations is calculated with respect to that change-point.

If $\tilde{T}_{\max,n}$ is larger than the pre-specified control limit, it signals a change and the estimated change-point is

$$\hat{\tau}_{\tilde{T}} = \text{argmax}_{1\leq k\leq n-1}|\tilde{T}_{k,n}|.$$

2.1.2 Correlation Structure

Hawkins (1977) states $Z_{\max,n}$ is the maximum of a set of correlated N(0,1) variables. The correlation between $Z_{k,n}$ and $Z_{m,n}$ $(1 \leq m < k \leq n-1)$ is $\rho_{mk} = \sqrt{\frac{m(n-k)}{k(n-m)}}$ under H_o, which makes the $Z_{1,n}, Z_{2,n}, \cdots, Z_{n-1,n}$ a Markovian process. That is one of the most important properties used in the follow-up discussion. The correlation between two adjacent $Z_{j,n}$ is $\rho_{j,j+1} = \sqrt{\frac{j(n-j-1)}{(j+1)(n-j)}}, j = 1, \cdots, n-2$. As a function of j/n, this correlation has an arced shape – close to 1 in the middle and approaching $1/\sqrt{2}$ at two ends if n is large. This high correlation among the $Z_{k,n}$ in the middle of the sequence and the relative independence of the $Z_{k,n}$ near the two ends means that the maximum $Z_{k,n}$ is more likely to occur near one end of the sequence than near the middle, so that in the null case $\hat{\tau}$ may be expected to have a "bath-tub" shaped distribution, symmetric about $n/2$ and rising steeply at the ends of the segments. This is indeed the case, as illustrated by calculation of the distribution of $\hat{\tau}$ in Hawkins (1977).

2.1.3 Asymptotic Results

Hawkins (1977) argued that $Z_{\max,n}$ should have an extreme value distribution when $n \to \infty$. It considered a Gaussian process $x(t), 0 < t < 1$ with mean 0, variance 1, and correlation function $\rho_{st} = \sqrt{s(1-t)/t(1-s)}$ for $s < t$. Then $Z_{1,n}, Z_{2,n}, \cdots, Z_{n-1,n}$ is a discrete version of the Gaussian process. If we define $\zeta(t) = t^{\frac{1}{2}}x[t/(1+t)], 0 < t < \infty$, $\zeta(t)$ will be a Brownian motion, and $x(t) < \lambda$ for all $t \in (0,1)$ is equivalent with

$\zeta(t) < \lambda\sqrt{t}$ for all $t \in (0, \infty)$. From the theory of Brownian motion, for any $\epsilon < 0$, $\zeta(t)$ exceeds $(1+\epsilon)\sqrt{2t \log|\log t|}$ at least once in every neighborhood of 0 and ∞, which means $|x(t)|$ will exceed $(1+\epsilon)\sqrt{2\log|\log t|}$ at least once in every neighborhood of 0 and ∞. Thus $Z_{\max,n}$ is unbounded. It goes to ∞ when $n \to \infty$. It is also discussed in this paper that although $Z_{\max,n}$ doesn't have an asymptotic distribution, it has an extreme value distribution which is plausibly $\sqrt{2\log\log n} + O(1)$.

Irvine (1982) gave a formal proof based on the Ornstein-Uhlenbeck process. Based on Theorem A.4.1 and Theorem 1.6.1 in Csörgő and Horváth (1997), a simple calculation shows that under H_o, for normal case, the $Z_{max,n}$ follows the extreme value distribution:

$$(2loglogn)^{-1/2} \max_{1 \le k < n} |Z_{k,n}| \xrightarrow{P} 1. \tag{2.1.9}$$

$$lim_{n\to\infty} P\left\{ A(logn) \max_{1 \le k < n} |Z_{k,n}| \le t + D(logn) \right\} = exp(-2e^{-t}). \tag{2.1.10}$$

where

$$A(x) = (2logx)^{1/2}, \qquad D(x) = 2logx + \frac{1}{2}loglogx - \frac{1}{2}log\pi. \tag{2.1.11}$$

Also

$$\hat{\tau}_Z / n \xrightarrow{D} \xi_0, \tag{2.1.12}$$

and $P(\xi_0 = 0) = P(\xi_0 = 1) = 1/2$.

The connection with the Brownian motion and the fact that this crosses any parabolic boundary infinitely often in every neighborhood of both zero and ∞ is what makes $Z_{\max,n}$ unbounded. Chopping off any region around zero and ∞ is enough to remove the unboundedness. James et al (1987) claimed that when we have n independent standard normal variables $X_1, X_2, \cdots, X_n$, if we define $S_k = x_1 + \cdots + x_k$, then for $1 \le n_0 < n_1 < n$ and $b > 0$,

$$P\Big(\max_{n_0 \le k \le n_1} \Big[\frac{kS_n/n - S_k}{\sigma\{k(1-k/n)\}^{\frac{1}{2}}} \Big] \ge b \Big) \sim 1 - \Phi(b) + b\phi(b) \int x^{-1} \nu(x + b^2/nx) dx, \tag{2.1.13}$$

where $\Phi(x)$ and $\phi(x)$ are cdf and pdf of standard normal distribution,

$$\nu(x) = 2x^{-2}exp[-2\sum_{1}^{\infty} n^{-1}\Phi(-xn^{1/2}/2)], \quad x > 0, \tag{2.1.14}$$

and the integral is on $(b(n_1^{-1} - n^{-1})^{\frac{1}{2}}, b(n_0^{-1} - n^{-1})^{\frac{1}{2}})$.

When n is large, the left-hand side of (2.1.13) can be approximated by Brownian motion $W(t)$ $(0 \leq t < \infty)$:

$$P(\max_{t_0 \leq t \leq t_1} W_0(t)/\{t(1-t)\}^{\frac{1}{2}} \geq b), \tag{2.1.15}$$

where $t_0 = n_0/n$, $t_1 = n_1/n$ and $W_0(t) = W(t) - tW(1)$ is a Brownian bridge on $[0, 1]$. When σ is unknown, similar results are present after we substitute sample standard deviation $\hat{\sigma}$ for σ.

It is easy to verify that $\frac{S_k - kS_n/n}{\sigma\{k(1-k/n)\}^{\frac{1}{2}}}$ is just $Z_{k,n}$ in (2.1.6) and the probabilities in (2.1.13) and (2.1.15) go to 0 when $b \to \infty$, which implies quarantining a fraction of the search interval at both ends will turn the extreme value distribution into an asymptotic distribution.

Mathematically, if we define

$$Z'_{\max,n} = \max_{[cn] \leq k \leq n-[cn]} |Z_{k,n}|, \tag{2.1.16}$$

where $c \in (0, \frac{1}{2})$, then $Z'_{\max,n}$ will have an asymptotic distribution.

2.2 Nonparametric Model

2.2.1 Existing Model

To adapt the change-point formulation to the nonparametric setting requires a nonparametric two-sample test statistic. Although not as substantial as for the normal case, the literature on the Phase I nonparametric change-point model is quite extensive.

Assume $X_1, X_2, \cdots, X_\tau, X_{\tau+1}, \cdots, X_n$ are independent, continuous random variables.

$$X_i \sim F(x), \quad \text{for } i = 1, 2, \cdots, \tau; \tag{2.2.1}$$

$$X_i \sim F(x+\theta), \quad \text{for } i = \tau+1, \cdots, n. \tag{2.2.2}$$

To test whether the process is in control, the hypothesis is

$$H_o : \theta = 0, \qquad H_a : \theta \neq 0.$$

Again, the change point τ is an unknown parameter, and the sample size n is a fixed number.

Pettitt (1979) proposed a U statistic based on the Mann-Whitney two-sample test. Let

$$D_{ij} = sgn(X_i - X_j) = \begin{cases} 1 & \text{if } X_i > X_j \\ 0 & \text{if } X_i = X_j \\ -1 & \text{if } X_i < X_j. \end{cases}$$

The U statistics $U_{k,n}$ is defined by D_{ij} as an anti-symmetric function.

$$U_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} D_{ij}, \quad 1 \leq k \leq n-1. \tag{2.2.3}$$

To test if there is a change, we use

$$U_{\max,n} = \max_{1 \leq k \leq n-1} |U_{k,n}|. \tag{2.2.4}$$

$$\hat{\tau}_U = \text{argmax}_{1 \leq k \leq n-1} |U_{k,n}|.$$

Let R_i be the rank within X_i with $X_1, X_2, \cdots, X_n$. From the two equivalent formulas of the Mann-Whitney statistic, it is easy to verify that

$$U_{k,n} = 2 \sum_{i=1}^{k} R_i - k(n+1).$$

It follows that

$$E(U_{k,n}) = 0, \quad Var(U_{k,n}) = \frac{k(n-k)(n+1)}{3}. \tag{2.2.5}$$

Note that the variance of $U_{k,n}$ depends on the putative split point k, being a maximum when $k = n/2$, and that $U_{\max,n}$ is the maximum of n random variables with different variance. This non-constant variance means that $U_{\max,n}$ has a strong tendency to split the sequence into roughly equal halves, a feature that makes it unsuitable for use in quality control work.

To avoid this problem of non-constant variance, Schechtman and Wolfe (1981) and (1984) proposed a weighted variant which standardizes the $U_{k,n}$ to constant variance. They defined

$$T_{k,n} = \frac{U_{k,n}}{\sqrt{k(n-k)(n+1)/3}}, \tag{2.2.6}$$

and

$$T_{\max,n} = \max_{1\le k\le n-1} |T_{k,n}|, \tag{2.2.7}$$

$$\hat{\tau}_T = \text{argmax}_{1\le k\le n-1}|T_{k,n}|.$$

From the asymptotic normality of Mann-Whitney statistics, $T_{k,n} \sim N(0,1)$ when k and $n-k$ both go to ∞.

2.2.2 Correlation Structure

It can be proved that when all H_0 holds, the correlation between $T_{m,n}$ and $T_{k,n}$ in nonparametric model is exactly the same as between $Z_{m,n}$ and $Z_{k,n}$ in parametric model.

THEOREM 2.2.1. *When H_o holds, the correlation between $T_{m,n}$ and $T_{k,n}$ is*

$$\tilde{\rho}_{mk} = corr(T_{m,n}, T_{k,n}) = \sqrt{\frac{m(n-k)}{k(n-m)}}.$$

Proof.

$$\begin{aligned} Cor(T_{m,n}, T_{k,n}) &= Cor\left(\tfrac{U_{m,n}}{m(n-m)(n+1)/3}, \tfrac{U_{k,n}}{k(n-k)(n+1)/3}\right) \\ &= Cor(U_{m,n}, U_{k,n}). \end{aligned}$$

$$\begin{aligned} Cov(U_{m,n}, U_{k,n}) &= E\left[(\sum_{i=1}^{k}\sum_{j=k+1}^{n} D_{ij})(\sum_{l=1}^{m}\sum_{h=m+1}^{n} D_{lh})\right] \\ &= \sum_{1\le i\le k,\ k+1\le j\le n,\ \ m\le l\le l,\ m+1\le h\le n} E(D_{ij}D_{lh}). \end{aligned} \tag{2.2.8}$$

If $i = l,\ j = h$:

$$E(D_{ij}D_{lh}) = E(D_{ij}^2) = P(X_i \neq X_j) = 1.$$

And there are $m(n-k)$ kinds of combinations for that. If $i = j, j \neq h$,

$$\begin{aligned} E(D_{ij}D_{lh}) =& P(X_i > X_j \text{ and } X_i > X_h) + P(X_i < X_j \text{ and } X_i < X_h) \\ & - P(X_i > X_j \text{ and } X_i < X_h) - P(X_i < X_j \text{ and } X_i > X_h). \end{aligned} \tag{2.2.9}$$

Define $p = (X_i > X_j > X_h)$,

$$P(X_i > X_j \text{ and } X_i > X_h) = P(X_i > X_j > X_n) + P(X_i > X_h > X_j) = 2p.$$

Similarly $P(X_i < X_j \text{ and } X_i < X_h) = 2p$. Then

$$E(D_{ij}D_{lh}) = 2p + 2p - p - p = 2p.$$

We know the sum of the four probabilities in equation (2.2.9) is the probability that any two of X_i, X_j, X_h are not the same, which should be 1 based on the continuous distribution assumption. So $6p = 1$ implies $p = 1/3$. Then

$$E(D_{ij}D_{lh}) = 1/3, \tag{2.2.10}$$

and there are $m(n-k)(n-m-1)$ kinds of combinations for that.

If $i \neq l, j = h$,

$$E(D_{ij}D_{lh}) = E(-D_{ji})(-D_{hl}) = 1/3, \tag{2.2.11}$$

and there are $(n-k)m(k-1)$ kinds of combinations for that.
If $h = i, j \leq l$,

$$E(D_{ij}D_{lh}) = E(D_{ij})(-D_{hl}) = -1/3, \tag{2.2.12}$$

and there are $m(k-m)(n-k)$ kinds of combinations for that.
If $h \neq l, i \neq h, j \leq l, j \leq h$,

$$E(D_{ij}D_{lh}) = E(D_{ij})E(-D_{hl}) = 0. \tag{2.2.13}$$

So combining all those, we have

$$\begin{aligned} Cov(U_{m,n}, U_{k,n}) =& m(n-k) + \frac{1}{3}m(n-k)(n-m-1) \\ &+ \frac{1}{3}(n-k)m(k-1) - \frac{1}{3}m(k-m)(n-k) \\ =& \frac{(1+n)m(n-k)}{3}. \end{aligned} \tag{2.2.14}$$

$$\begin{aligned} \tilde{\rho}_{mk} &= Cor(T_{m,n}, T_{k,n}) = Cor(U_{m,n}, U_{k,n}) \\ &= \frac{(1+n)m(n-k)/3}{[(1+n)m(n-m)/3]^{1/2}[(1+n)k(n-k)/3]^{1/2}} \\ &= \sqrt{\frac{m(n-k)}{k(n-m)}}. \end{aligned} \tag{2.2.15}$$

This concludes the proof. □

This is exactly the same as normal case. That the parametric $Z_{k,n}$ and nonparametric $T_{k,n}$ have the same correlation structure and, for moderately large k and $n-k$ approximately the same $N(0,1)$ distribution might lead one to expect that $Z_{\max,n}$ and $T_{\max,n}$ would have similar distributions for sufficiently large n, as would $\hat{\tau}_Z$ and $\hat{\tau}_T$, but this turns out not to be the case. The reason lies in the non-normality of $T_{k,n}$ when k or $n-k$ is small, which is precisely the region in which $\hat{\tau}_Z$ is likely to lie.

2.2.3 Asymptotic Results

Csörgő and Horváth (1997) also gave asymptotic results for U statistics with anti-symmetric kernel function, including our Mann-Whitney based statistic as a special case. From their Theorem 2.4.12 and Theorem 2.4.14, we could easily get the asymptotic distribution for $T_{max,n}$.

THEOREM 2.2.2.

$$lim_{n\to\infty}P\left\{A(logn)\max_{1\le k<n}|T_{k,n}|\le t+D(logn)\right\}=exp(-2e^{-t}). \tag{2.2.16}$$

where $A(x)$ and $D(x)$ are defined in (2.1.11)

Also

$$\hat{\tau}_T/n\xrightarrow{D}\xi_0, \tag{2.2.17}$$

and $P(\xi_0=0)=P(\xi_0=1)=1/2$.

Proof. Let anti-symmetric kernel function $h(x,y)=sgn(x-y)=-h(y,x)$. Now $U_{k,n}=\sum_{i=1}^{k}\sum_{j=k+1}^{n}h(X_i,X_j)$. Then it is easy to see that $Eh(X_1,X_2)=0$ and $Eh^2(X_1,X_2)<\infty$. Assume the cdf of X_i is $F(\cdot)$, then $\tilde{h}(t)=Eh(X_1,t)=P(X_1>t)-P(X_1<t)=1-2F(t)$. We know $F(X_1)\sim U(0,1)$,$1-2F(X_1)\sim U(-1,1)$, so

$$E(1-2F(X_1))^2=\int_{-1}^{1}\frac{1}{2}x^2dx=1/3.$$

So $0<E(\tilde{h}^2(X_1))<\infty$. Let $\sigma=\sqrt{E(\tilde{h}^2(X_1))}$, Theorem 2.4.12 in Csörgő and Horváth(1997) says

$$lim_{n\to\infty}P\left\{A(logn)\max_{1\le k<n}\frac{|U_{k,n}|}{\sigma(k(n-k+1)n)^(1/2)}\le t+D(logn)\right\}=exp(-2e^{-t}),$$

which is just simply

$$lim_{n\to\infty}P\left\{A(logn)\max_{1\le k<n}|T_{k,n}|\le t+D(logn)\right\}=exp(-2e^{-t}).$$

□

This is the same asymptotic distribution quoted earlier for the parametric case. However in view of the very slow convergence, this does not necessarily mean that the behavior of the two statistics is the same for moderate sample sizes. This question will be investigated by simulation.

2.3 Simulation study

2.3.1 Simulation for $n = 20$

A simulation study is performed to investigate the results of section 2.1.3 and 2.2.3. Figure 2.1 is for phase I analysis with sample size equal to 20. It is based on 50000 runs. For each run, $n = 20$ independent normal distributed random variables are generated, and $Z_{k,n}$, $U_{k,n}$, $T_{k,n}$ are calculated based on these 20 observation.

Figure 2.1 shows the shape of probability mass function of $\hat{\tau}_Z, \hat{\tau}_U$ and $\hat{\tau}_T$ under H_o. As expected, there is a clearly symmetric bathtub shape for $Z_{k,n}$ and $T_{k,n}$. They are close to each other, except at the edges where $k = 1$ and $n - 1$. $T_{k,n}$ is less likely to give a signal at the very edge where $k = 1, n - 1$, but slightly more likely at $k = 2, 3, n - 3, n - 2$. The difference between the two decreases as k goes from edges to middle. That means $T_{k,n}$ is less inclined to separate a single observation.

When we consider the Phase II analysis in future, a good chart should give a signal as soon as possible. It implies a large segment on the left of change-point and small segment on the right. This bathtub shape suggests charts based on $T_{k,n}$ should be nice, just as $Z_{k,n}$ in normal case. Instead, $U_{k,n}$ has a arc-shaped distribution, which tends to give a signal in the middle. That is because the variance for $U_{k,n}$ is much larger in the middle than at the edge. So it is more likely to have an extreme value in the middle. It will definitely do a better job when the change actually happens in the middle. But in general, it may not be appropriate in SPC, especially under the

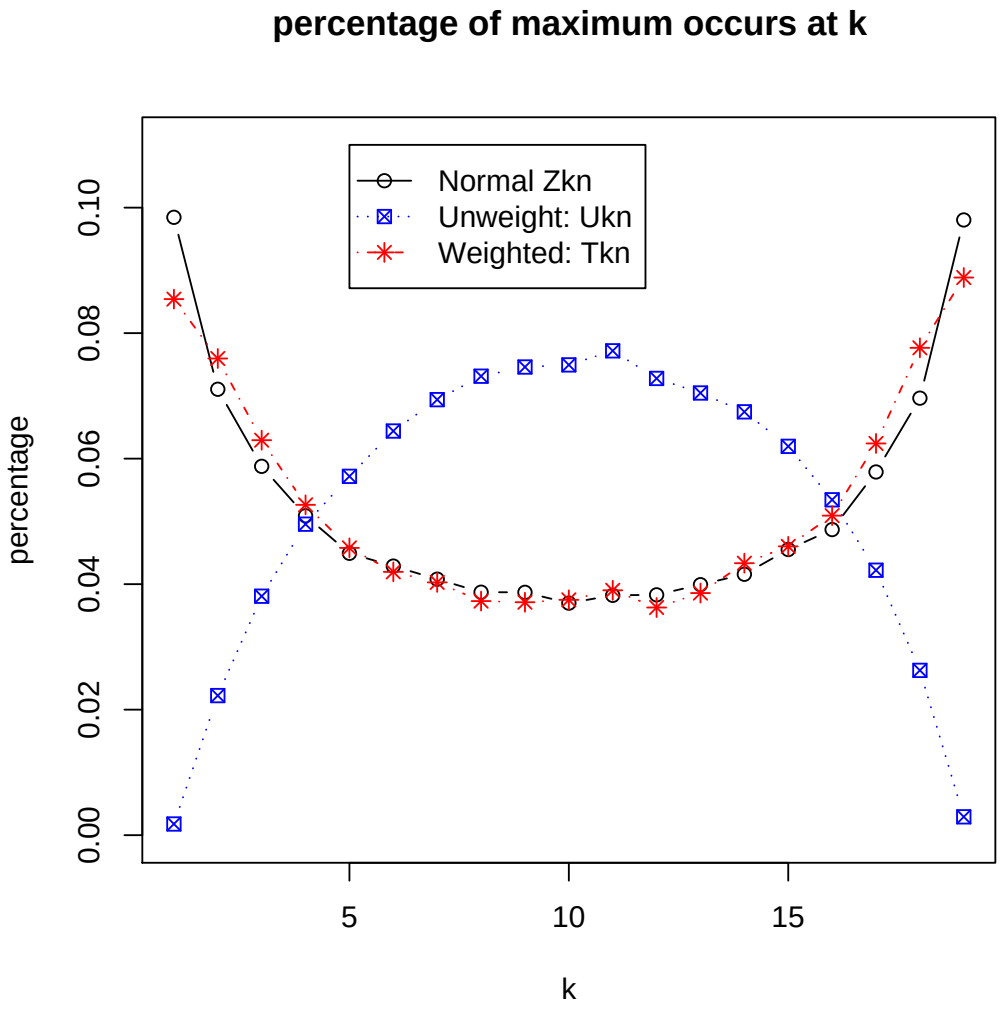

Figure 2.1: Probability mass function estimation for $\hat{\tau}_Z, \hat{\tau}_U$ and $\hat{\tau}_T$ under H_o estimated by 50000 simulation, when underlying distribution is normal

Phase II setting. So in the following discussion we only consider $T_{k,n}$.

Figure 2.2 shows the situation for normal(0,1), t(3), exponential(3) and cauchy(0,1) respectively. You can see behavior of $T_{k,n}$ stays the same for different distributions, while $Z_{k,n}$ varies a lot. It is determined by the distribution-free property of Mann-Whitney statistics.

2.3.2 When n varies

Another simulation was carried out to investigate the effect of n. For $n = 30, 50, 100, 500$, the same simulation of size 50000 is carried out, but only $Z_{k,n}$ and $T_{k,n}$ are kept. To make results more comparable, $\hat{\tau}_Z$ and $\hat{\tau}_T$ are rescaled by $1/n$. Result is in Figure 2.3.

We can see that the gap between the two for $k = 1, n-1$ doesn't disappear when n gets larger. The probabilities of maximum occurring at $k = 1$ or $n-1$ for $T_{k,n}$ keep staying well below $Z_{k,n}$. For other k, those two seems to be close when n is large.

On the other hand, the percent for maximums to occur at $k = 1$ or $n-1$ decreases when n increases. For $n = 30$, there are about 13% maximums of $T_{k,n}$ fall at either $k = 1$ or $n - 1$. It drops to 9% for $n = 50$, 5% for $n = 100$ and 2% for $n = 500$.

Actually, provided k and $n - k$ are large, the approximate normality of the $T_{k,m}$ and identical correlation to the parametric case mean that the two will behave much the same. The only thing that will unravel the equivalence is if there is a strong tendency to split off so few observations that we don't yet have decent normality. But that does not seem to happen if n is large.

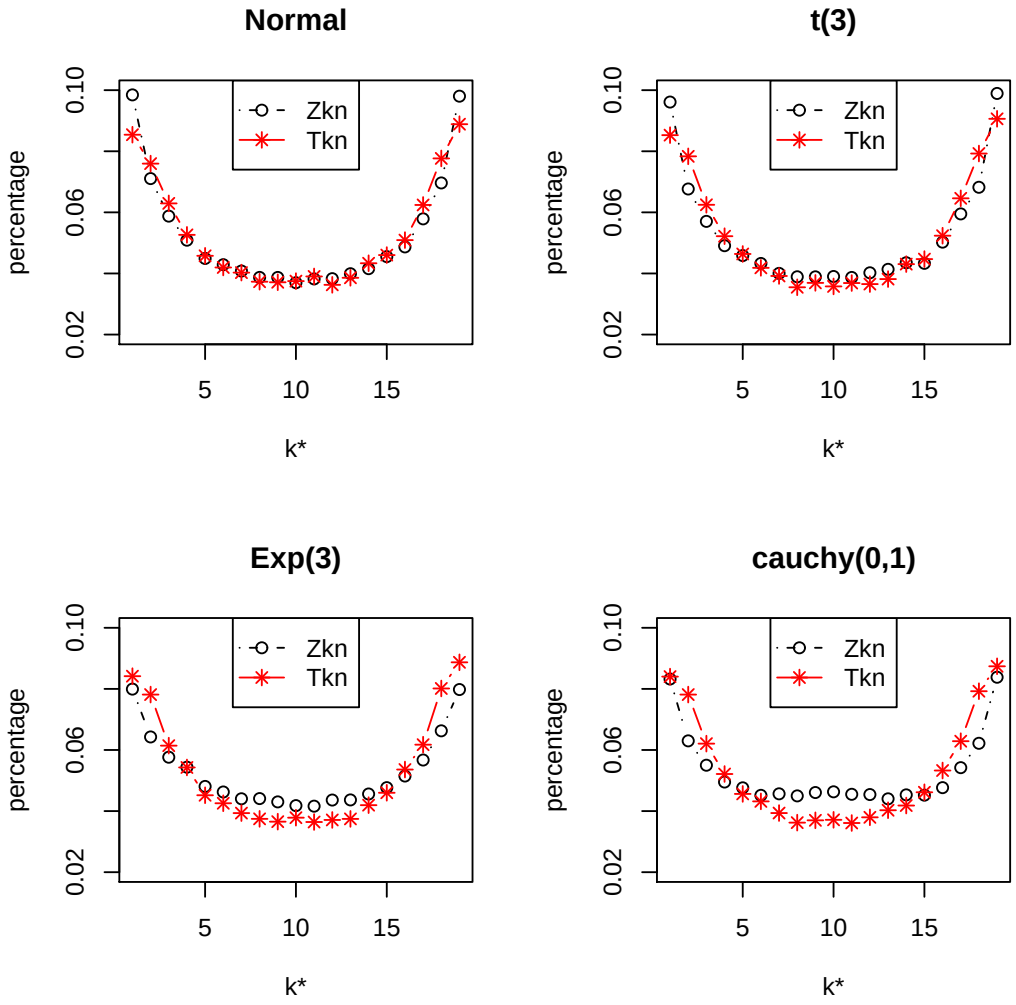

Figure 2.2: Distribution for $\hat{\tau}_Z$ and $\hat{\tau}_T$ under H_o when underlying distribution is Normal(0,1), t(3), exponential(3) and cauchy(0,1)

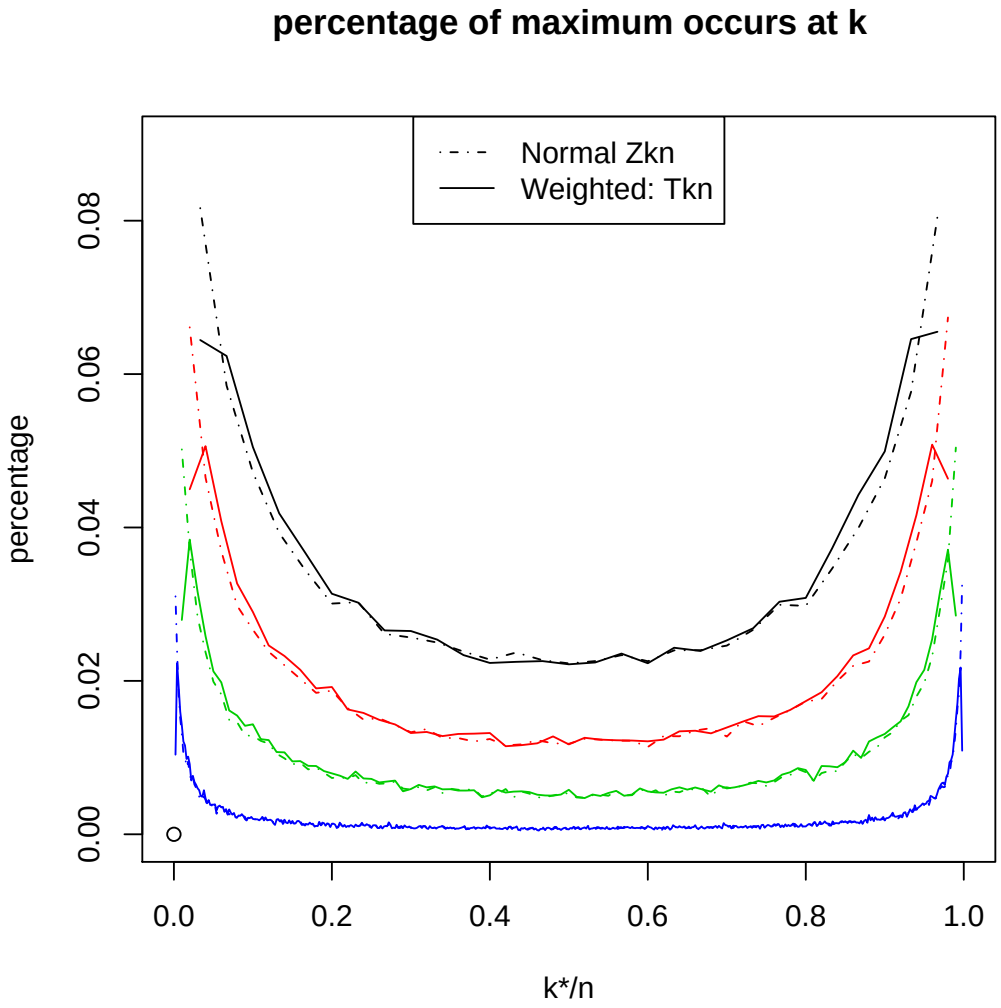

Figure 2.3: Probability mass function estimation of $\hat{\tau}_Z$ and $\hat{\tau}_T$ for $n = 30, 50, 100, 500$ (from top to bottom) under H_o, when underlying distribution is normal

2.3.3 Quality of normal approximation

Ideally, we should use the exact distribution for $T_{max,n}$ to set up control limits. However, this is computationally intolerable. Schechtman (1982) discussed the null distribution of $T_{max,n}$ as well as its power, but used a Monte Carlo method, which limits its generality. In this section, we investigate, at a more fundamental level, the quality of the normal approximation. We could get the exact distribution of these $T_{k,n}$ using theory, but when comparing $T_{max,n}$ with normal approximation later on, we need to rely on simulation. So right now, for investigation purpose, we just use simulation.

Based on the same simulation as last section, we get the cumulative density function(cdf) of $T_{k,n}$, and compare it to the normal approximations. Figure 2.4 is for $k = 1$. Here we notice an obvious deviation from normal even for n as large as 100. From the trend, it seems that this discrepancy is not going to disappear as n grows. Fortunately, the problem seems negligible for larger k. This is showed in Figure 2.5. Even for $k = 2, n = 10$, the normal approximation is good, and improves as k or n increases.

Figure 2.6 shows the cdf for $Z_{max,n}$ and $T_{max,n}$. Notice for $k = 10$, the main difference comes from the abnormal jump as estimated from $50,000$ sequences at $T_{max,n} = 1.57$, which corresponds to the highest possible value of $|T_{1,10}|$. However, the tail probabilities are getting closer as n increases. When $n = 100$, the difference between the two is very tiny. These results agree with what we have found before: The approximation of $T_{\max,n}$ by $Z_{\max,n}$ seems reasonably good for large n, but not appropriate for small n, mainly due to the non-normality when $k = 1$.

In summary of phase I analysis, it is found that parametric and nonparametric change-point models are similar in many aspects. The parametric $Z_{k,n}$ and nonparametric $T_{k,n}$ have identical correlation and approximately the same standard normal

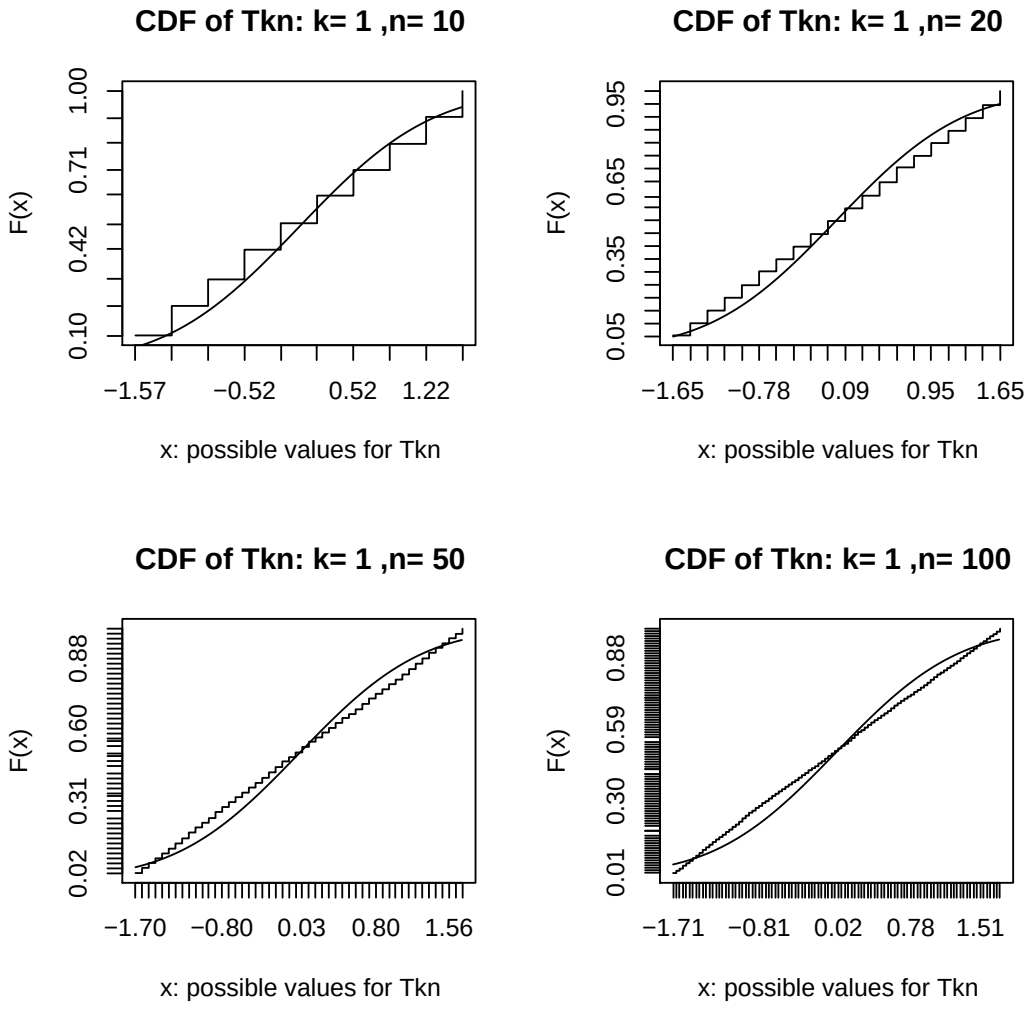

Figure 2.4: CDF of $T_{k,n}$ estimated by simulation and normal approximation for $k = 1$

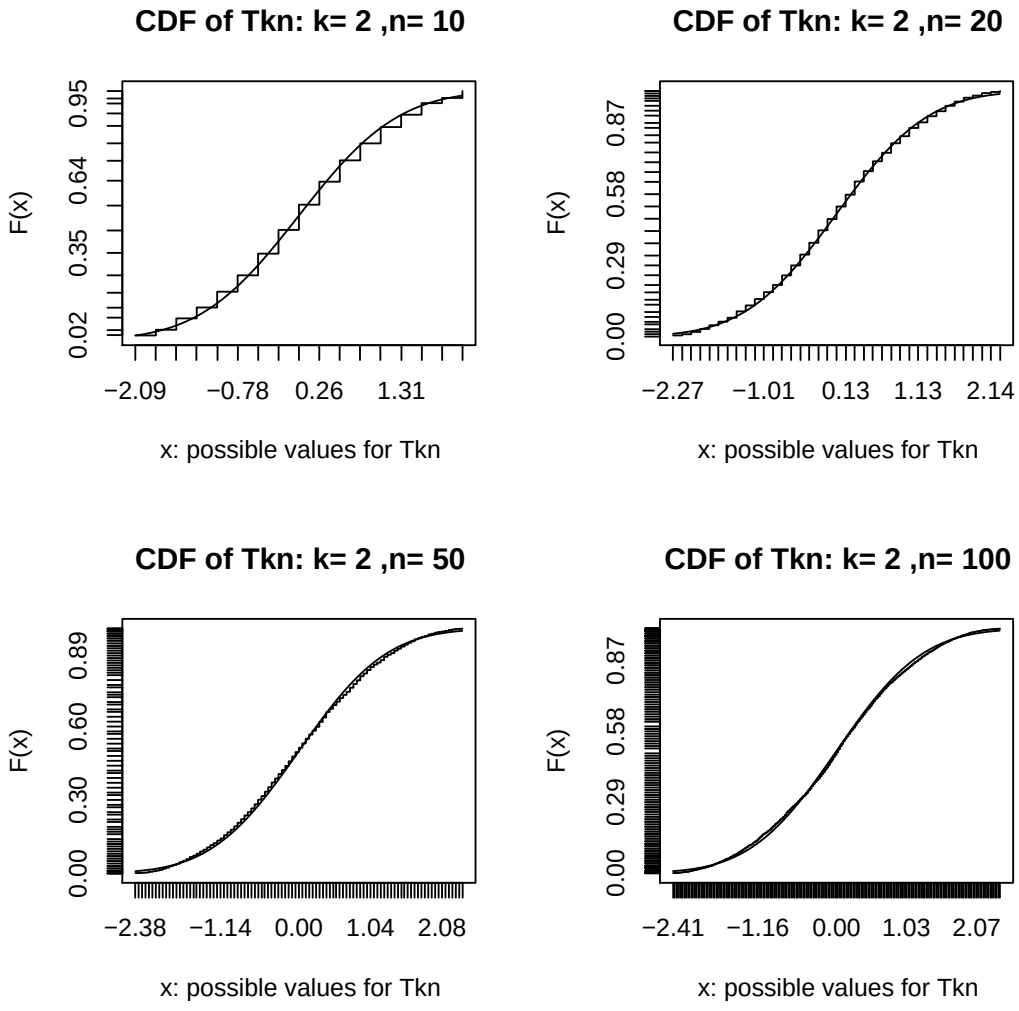

Figure 2.5: CDF of $T_{k,n}$ estimated by simulation and normal approximation for $k = 2$

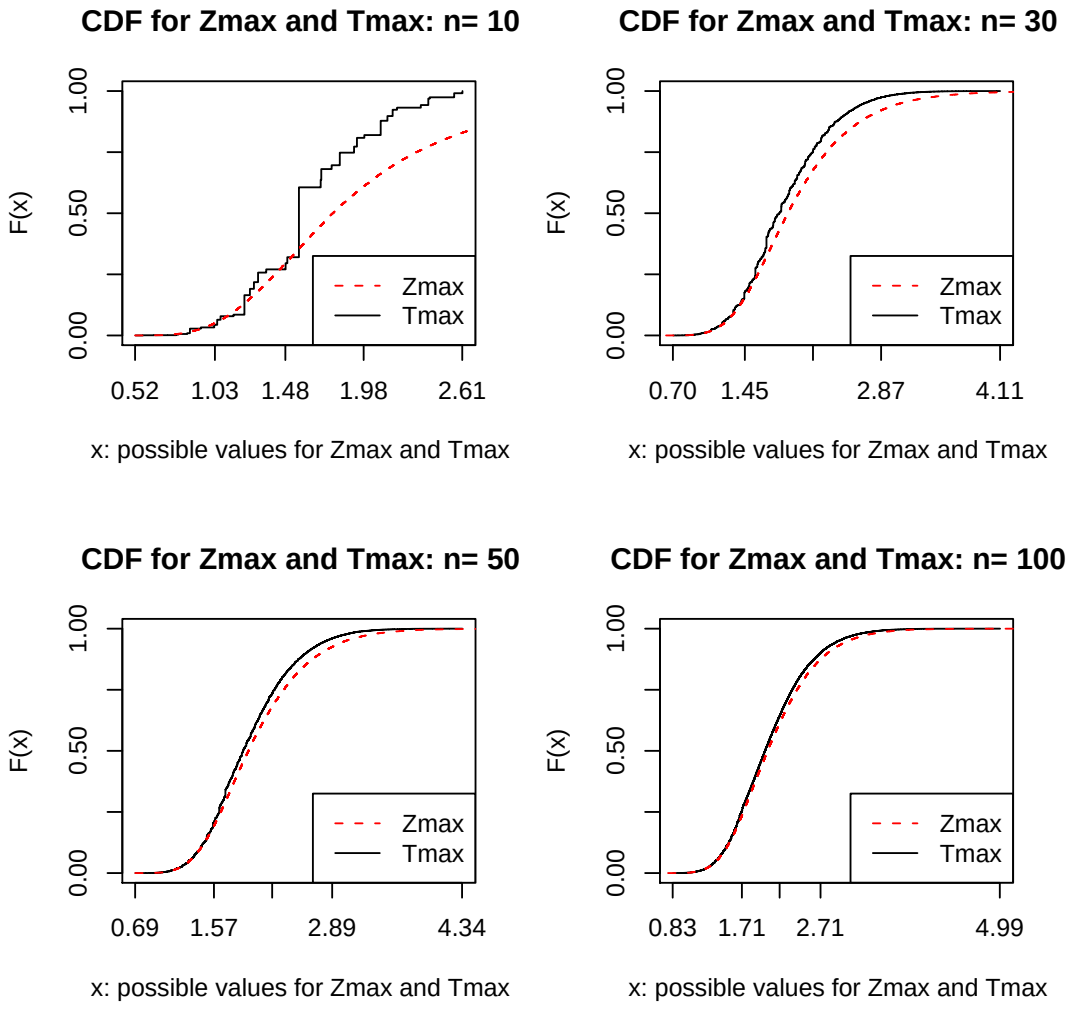

Figure 2.6: CDF of $Z_{max,n}$ and $T_{\max,n}$ estimated by simulation

distribution for moderately large k and $n-k$. These common characteristics give them similar "bath-tub" shaped distributions for $\hat{\tau}$. However, the two distributions are not identical: the parametric is more likely to split off a short segment at either end. The different behaviors are due to the non-normality of $T_{k,n}$ when k or $n-k$ is small. Fortunately, the discrepancy doesn't seem to affect the equivalence of asymptotic results. $Z_{max,n}$ and $T_{max,n}$ are proved to have identical extreme value distribution. It is also very interesting that chopping off a fraction of test statistic at both end will be enough to turn the extreme value distribution into an asymptotic distribution.

Chapter 3

The Parametric Change-point Model in Phase II Analysis

3.1 CUSUM Chart

There are several classical charts used in SPC: Shewhart chart by Shewhart (1925), Cumulative Sum chart by Page (1954) and Exponentially Weighted Moving Averages chart by Roberts (1959).

The Cumulative Sum chart (CUSUM chart) is the most relevant one for phase II change-point problems. It is well known that Page's CUSUM chart is essentially maximizing the likelihood to estimate the unknown change-point. Assume $X_1, X_2, \cdots, X_\tau, X_{\tau+1}, \cdots, X_n, \cdots$ are independent random variables. The density function is f_o for $X_1, \cdots, X_\tau$, and is f_1 for $X_{\tau+1}, \cdots, X_n, \cdots$. Let N be the stopping time, then

$$N_{CUSUM} = \inf\{n : \max_{0 \le k \le n-1} \sum_{i=k+1}^{n} \log(f_1(X_i)/f_o(X_i)) \ge c_\alpha\}, \tag{3.1.1}$$

where c_γ is the appropriate control limit which leads to an in-control ARL of $1/\alpha$.

The change-point estimate is

$$\hat{\tau}_{CUSUM} = \text{argmax}_{0\le k\le N_{CUSUM}-1} \sum_{i=k+1}^{N_{CUSUM}} \log(f_1(X_i)/f_o(X_i)).$$

And it is proved that CUSUM chart is optimal in the sense of minimax property.

The CUSUM rule (3.1.1) is very convenient in practice, since it has the simple update algorithm:

$$N_{CUSUM} = \inf\{n : l_n \ge c_\alpha\}, \tag{3.1.2}$$

where $l_0 = 0$ and

$$l_n = \{l_{n-1} + \log(f_1(X_n)/f_o(X_n))\}^+. \tag{3.1.3}$$

However, the CUSUM rule relies on a strong assumption that the densities before and after the shift are known, i.e. $f_1(\cdot)$ and $f_o(\cdots)$ are known. This is quite unrealistic in practice. In quality control, there are all kinds of causes that can lead to a change of distribution. In general, there is no particular fixed distribution that the process will shift to.

In Barnard (1959), a CUSUM rule, also known as generalized Likelihood ratio (GLR) rule, is proposed for the parametric family with known pre-change parameter and unknown post-change parameter:

$$N_G = \inf\{n : \max_{0\le k\le n-1} \sup_{\theta\in\Theta} \sum_{i=k+1}^{n} \log \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)} \ge c_\alpha\}, \tag{3.1.4}$$

and

$$\hat{\tau}_G = \text{argmax}_{0\le k\le N_G-1} \sup_{\theta\in\Theta} \sum_{i=k+1}^{N_G} \log \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)}.$$

Specially, when the distribution is normal with known variance, the generalized CUSUM rule is

$$N_G = \inf\{n : \max_{0\le k\le n-1} (X_{k+1} + \cdots + X_n)^2/2(n-k) \ge c_\alpha\}. \tag{3.1.5}$$

This generalized CUSUM lowers the bar of the strong assumptions in original CUSUM chart and is more into reality. However, the cost is the computational and theoretical simplicity. The simple update algorithm of (3.1.3) doesn't hold any more in the generalized CUSUM. Considering the facilities available in those days, the window-limited approach was used to reduce the computational burden. This will be discussed in section 6.1. In addition, the relationship between the ARL and the control limit c_α remained a headache for a long time, until it was solved by Siegmund and Venkatraman (1995). It will be discussed in Chapter 5.

3.2 HQK – The Parametric Change-Point Model in Phase II Analysis

HQK is a change-point control chart proposed by Hawkins, Qiu and Kang (2003) for normal distribution. As each new observation accrues, the change-point control chart looks at every previous time as a potential change point, carrying out a two-sample t test between the observations preceding, and those following that time. If the maximum of these t statistics exceeds a specified control limit, the method signals that a change has occurred, and estimates the time of change as the maximizing split point. In mathematical notation, when variance σ is known,

$$N_{HQK} = \inf\{n : \max_{1\le k\le n-1} |Z_{k,n}| \ge \tilde{h}_{\alpha,n}\}, \tag{3.2.1}$$

and

$$\hat{\tau}_{HQK} = \text{argmax}_{1\le k\le N_{HQK}-1}|Z_{k,N_{HQK}}|,$$

where $Z_{k,n}$ is defined in (2.1.6) and is basically the two sample Z test statistic between left and right sections of the observations is calculated with respect to that change-point.

When variance σ is unknown,

$$N_{HQK} = \inf\{n : \max_{1\le k\le n-1} |\tilde{T}_{k,n}| \ge \tilde{h}_{\alpha,n}\}. \tag{3.2.2}$$

where $\tilde{T}_{k,n}$ is defined in (2.1.8) and is basically the two sample T test statistic between left and right sections of the observations is calculated with respect to that change-point.

3.3 Connection between CUSUM and HQK

CUSUM, generalized CUSUM and HQK are all based on likelihood ratio tests. Assume $X_1, X_2, \cdots, X_\tau, X_{\tau+1}, \cdots, X_n, \cdots$ are independent normal random variables. When θ_o and θ are both known,

$$\begin{aligned} -\log(\text{Likelihood Ratio}) &= \log \frac{\max_{0\le k\le n-1}[\Pi_{i=1}^k f_{\theta_o}(X_i)\cdot \Pi_{i=k+1}^n f_\theta(X_i)]}{\Pi_{i=1}^n f_{\theta_o}(X_i)} \\ &= \log \max_{0\le k\le n-1} \Pi_{i=k+1}^n \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)} \\ &= \max_{0\le k\le n-1} \sum_{i=k+1}^n \log \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)}. \end{aligned}$$

This is the CUSUM rule.

Similarly, when θ_o is known and θ is unknown,

$$\begin{aligned} -\log(\text{Likelihood Ratio}) &= \log \frac{\max_{0\le k\le n-1} \sup_{\theta\in\Theta}[\Pi_{i=1}^k f_{\theta_o}(X_i)\cdot \Pi_{i=k+1}^n f_\theta(X_i)]}{\Pi_{i=1}^n f_{\theta_o}(X_i)} \\ &= \max_{0\le k\le n-1} \sup_{\theta\in\Theta} \sum_{i=k+1}^n \log \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)}. \end{aligned}$$

This is the generalized CUSUM rule.

When θ_o and θ are both unknown,

$$\begin{aligned}
-\log(\text{Likelihood Ratio}) &= \log \frac{\max_{1\le k\le n-1} \sup_{\theta\in\Theta,\theta_o\in\Theta}[\Pi_{i=1}^{k} f_{\theta_o}(X_i)\cdot\Pi_{i=k+1}^{n} f_{\theta}(X_i)]}{\sup_{\theta_o\in\Theta}\Pi_{i=1}^{n} f_{\theta_o}(X_i)} \\
&= \max_{1\le k\le n-1}[-\sum_{i=1}^{k}\frac{(X_i-\bar{X}_{0,k})^2}{2\sigma^2}-\sum_{i=k+1}^{n}\frac{(X_i-\bar{X}_{k,n})^2}{2\sigma^2}\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{2\sigma^2}] \\
&= \frac{1}{2\sigma^2}\max_{1\le k\le n-1}[k\bar{X}_{0,k}^2+(n-k)\bar{X}_{k,n}^2-n\bar{X}^2] \\
&= \frac{1}{2\sigma^2}\max_{1\le k\le n-1}[k\bar{X}_{0,k}^2+(n-k)\bar{X}_{k,n}^2-\frac{(k\bar{X}_{0,k}+(n-k)\bar{X}_{k,n})^2}{n}] \\
&= \max_{1\le k\le n-1}[\frac{1}{2}\frac{k(n-k)}{n}\frac{(\bar{X}_{0,k}-\bar{X}_{k,n})^2}{\sigma^2}] \\
&= \max_{1\le k\le n-1} Z_{k,n}^2/2,
\end{aligned}$$

where $\bar{X}_{0,k}$, $\bar{X}_{k,n}$ and $Z_{k,n}$ are defined in (2.1.3), (2.1.4) and (2.1.6) respectively. $\bar{X}$ is the sample mean of all the observations. Now it is equivalent with using $|Z_{k,n}|$, which is the HQK rule.

So essentially,

$$N_{HQK}=\inf\{n:\max_{1\le k\le n-1}\log\frac{\sup_{\theta\in\Theta,\theta_o\in\Theta}[\Pi_{i=1}^{k} f_{\theta_o}(X_i)\cdot\Pi_{i=k+1}^{n} f_{\theta}(X_i)]}{\sup_{\theta_o\in\Theta}\Pi_{i=1}^{n} f_{\theta_o}(X_i)}\ge\tilde{h}_{\alpha,n}\}. \tag{3.3.1}$$

From here, we can see HQK is more generalized than the generalized CUSUM. Besides the unknown parameter in post-change distribution, change-point model can handle the case where pre-change distribution parameter is also unknown.

It should also be noted that:

1. The possible change-points in HQK are from 1 to n, where the ones in CUSUM and generalized CUSUM are from 0 to n. Because in CUSUM and generalized CUSUM, the in-control parameter is known and fixed to be θ_o, the shift from 1st observation can be identified. However, in HQK, θ_o is unknown, so we need at least one observation to represent the in-control case.

2. The control limit in HQK is a function of current sample size n. It varies when new observation comes in. Actually, it is designed to keep the conditional probability of a false alarm at α at anytime, given that there was no false alarm in all the previous tests. But in CUSUM and generalized CUSUM, the control limit only depends on the in-control ARL, and stays the same all the time.

Chapter 4

The Nonparametric change-point Model in Phase II Analysis

4.1 Model Set-Up

Assume $X_1, X_2, \cdots, X_\tau, X_{\tau+1}, \cdots$ are independent, continuous random variables.

$$X_i \sim F(x), \quad \text{for } i = 1, 2, \cdots, \tau; \tag{4.1.1}$$

$$X_i \sim F(x+\theta), \quad \text{for } i = \tau+1, \tau+2, \cdots \tag{4.1.2}$$

To test whether the process is in control, the hypothesis is

$$H_o : \theta = 0, \qquad H_a : \theta \neq 0.$$

Again, the change point τ is an unknown parameter. But the sample size keeps increasing when new observation comes in, since it is for an ongoing phase II analysis.

Paralleling the HQK methodology, as each new observation accrues, we compute the $T_{\max,n}$, which is defined in (2.2.7). If this exceeds a cutoff $h_{\alpha,n}$, then we conclude that a change has occurred, and estimate the change-point from the maximizing k.

Else we conclude the process is still IC and move on to the next observation.
Then our decision rule can be writen as:

$$N_{NPC} = \inf\{n : \max_{1\leq k\leq n-1} |T_{k,n}| \geq h_{\alpha,n}\}, \tag{4.1.3}$$

and

$$\hat{\tau}_{NPC} = \text{argmax}_{1\leq k\leq N_{NPC}-1}|T_{k,N_{NPC}}|,$$

where $T_{k,n}$ is defined in (2.2.6).

Basically, when a new observation accrues, we look at every previous time as a potential change point, carrying out a two-sample Mann-Whitney test between the observations preceding, and those following that time. If the maximum of these Mann-Whitney statistics exceeds the control limit $h_{\alpha,n}$, the method signals that a change has occurred.

Similar with parametric HQK setting, $h_{\alpha,n}$ is chosen so that the conditional probability of a false alarm at any observation, given that there was no false alarm at all the previous test, is a constant α. And this will be discussed in the next section.

4.2 Computation and Complexity

The parametric HQK algorithm requires storage of a table of running sums and sums of squared deviations – $2n$ words in all. Using computationally efficient numerically stable updates of these summary tables is fast, and as n increases, the only significant computation is the search across previous values for the location of the maximizing change point. This uses $12n$ arithmetic operations and so scales linearly.

Turning to the nonparametric case, at the first glance, the computational issue seems more intractable, because calculating the rank sum requires ordering the data, an $O(nlogn)$ operation, while the definition form of $U_{k,n}$ is $O(n^2)$. In the ongoing

analysis with new data come in continuously, all orderings change as each new observation is added and the rankings need continual updating.

More careful analysis however shows that the Mann-Whitney statistic has a simple update:

for $1 \leq k \leq n$

$$U_{k,n+1} = \sum_{i=1}^{k} \sum_{j=k+1}^{n+1} D_{ij} = \sum_{i=1}^{k} \left[\sum_{j=k+1}^{n} D_{ij} + D_{i,n+1} \right] = U_{k,n} + \sum_{i=1}^{k} D_{i,n+1}. \qquad (4.2.1)$$

Thus each $U_{k,n+1}$ can be calculated from $U_{k,n}$ using just two additions – one to update the running total $\sum_{i=1}^{k} D_{i,n+1} = \sum_{i=1}^{k-1} D_{i,n+1} + D_{k,n+1}$ and one to add this running total to $U_{k,n}$.

Computing $T_{k,n}$ from $U_{k,n}$ requires an additional four operations, making a total of $6n$ operations, leading to the surprising conclusion that the nonparametric formulation also scales linearly, and is computationally faster than the parametric.

The next major hurdle is setting the sequence of control limit $h_{\alpha,n}$. As in the parametric HQK setting, we fix the conditional probability of a false alarm at any observation, given that there was no false alarm at all the previous test. The ideal $h_{\alpha,n}$ should be set such that this conditional probability of a false alarm at any n is a constant α. This depends on the null distribution of $T_{\max,n}$, which is based on $T_{k,n}$. However, the exact distribution is discrete and varies as n or k changes, and requires finding out all the possible arrangements of the ranks. That will be a nightmare when n is large as discussed in Schechtman (1982), and this problem is even more severe in Phase II study. We will attack it by large-scale simulation.

In principle, it is possible to start testing for a change from the third observation, since the change-point model does not rely on parameter estimates. However, because of the discrete nature of the distribution of $T_{\max,n}$, we need to have some minimum number of "warm-up" cases before we start the testing, so that it becomes possible to get the sort of small α needed in SPC.

A warm-up of 14 cases allows α values as small as 1/3432, which is small enough for SPC use. With 14 warm-up data, the method starts actual monitoring from the 15th process reading.

The version of the nonparametric changepoint procedure that was studied used this approach - 14 initial observations are accumulated without any testing, and the testing starts on the fifteenth. Forty million sequences of length 1000 were simulated to find cutpoints up to $n = 1000$. These cutpoints are presented in Table 4.1, the columns corresponding to in-control average run lengths of 50, 100, 200, 500, 1000 and 2000 respectively. Not all n values are listed explicitly; you can carry entries forward. For example, there is an entry for $n = 60$ and one for $n = 70$; use the $n = 60$ entry for all values between 60 and 70.

The table suggests that the cutpoints stabilize to constant values after $n = 300$. From the discussion in section 2.1.3 and 2.2.3, we saw that the test statistic $Z_{max,n}$ for parametric and $T_{max,n}$ for nonparametric in phase I analysis both have extreme value distribution, which implies that the control limit, i.e. 95% quantile will go to ∞ with probability one when $n \to \infty$. However, it is surprising that in phase II analysis, both control limit stabilize when $n \to \infty$. This is due to that phase II model uses the conditional probability of a false alarm at any observation, given that there was no false alarm at all the previous test. So it only looks at the sequences not having any signals before, which substantially lowers the value of test statistic.

4.3 Example

Aluminum smelters monitor the composition of the feed material to respond to changes in composition. One of the nuisance constituents is silica. Table 4.2 lists a data set (kindly provided by Len Homer) on the silica concentrations in percent of the feed to a smelter. A glance at the data is enough to show that the numbers do

not follow a normal distribution, leading to the possibilities of transforming the data to normality, or using a nonparametric procedure. Two problems with the transformation route are that the sample size is quite limited, and the fact that if there is a shift somewhere in the sequence, then the data are not from a single distribution but are a mixture of two distributions, which confounds the attempt to find a suitable transformation.

Despite these reservations, a log transformation seems reasonably successful in getting the data to normality, and also accords with the general experience that low-concentration constituents tend to follow approximately lognormal distributions. Based on this, we analyze the sequence of natural-log-transformed silica values using the parametric HQK approach.

The nonparametric change-point model can be applied immediately to the data on the natural scale.

The results are shown in Figure 4.1 and Figure 4.2 respectively. In each figure, the left panel shows the change-point statistic $Z_{max,n}$ or $T_{\max,n}$ along with the control limit for an in-control ARL of 500. The right panel shows the maximizing k - the estimate of the last in-control reading.

Both nonparametric and parametric control charts conclude that there is a location shift midway through the series. The nonparametric chart exceeds its control limit at $n = 37$ and remains above the control limit to the end of the data. The parametric chart signals a bit later – at $n = 39$, but then dips back below the control limit before emerging at process reading $n = 44$.

As for the estimate of the change point, both methods vacillate initially between $\hat{\tau} = 31$ and $\hat{\tau} = 28$ before settling on the former. The nonparametric approach is, however, faster to make up its mind.

In this example then, the nonparametric change-point control chart provides results that are compatible with those of the parametric method, but is arguably more

decisive. These apparent advantages come on top of the fact that there is no need to struggle with searching for the appropriate transformation and worrying about the distribution issue at all.

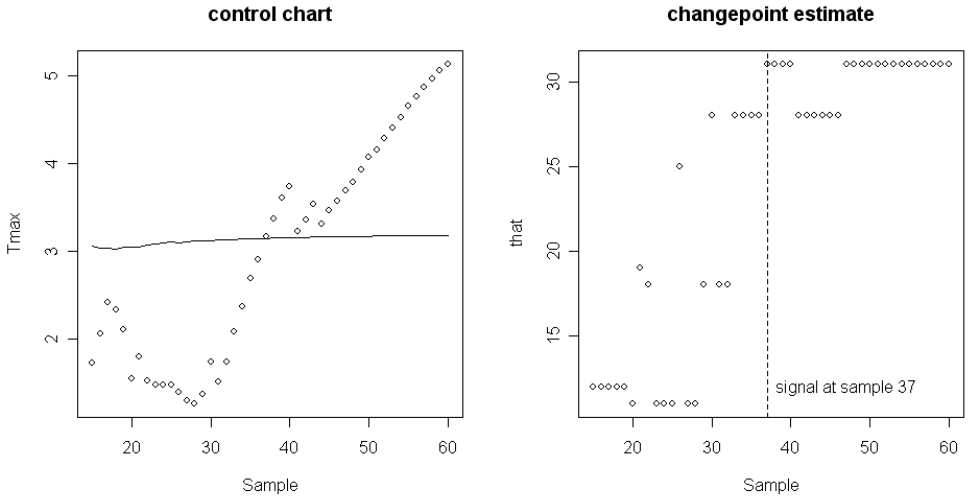

Figure 4.1: Nonparametric changepoint model on original SiO2 data: $T_{\max,n}$ and $h_{n,0.002}$

4.4 Evaluation of Performance

The performance of the chart is measured in terms of its ARL following a step change in distribution. Apart from the in-control ARL, two factors influence the performance – the magnitude of the shift, and the length of time for which the process runs in control before the step change. This second factor may be unintuitive initially, but as the change point formulation involves two-sample comparisons, its performance is affected by the size of the in-control data set as well as the run of out-of-control data. In addition, it is of interest to know how the performance of the nonparametric chart compares with that of the parametric HQK chart in circumstances where both would

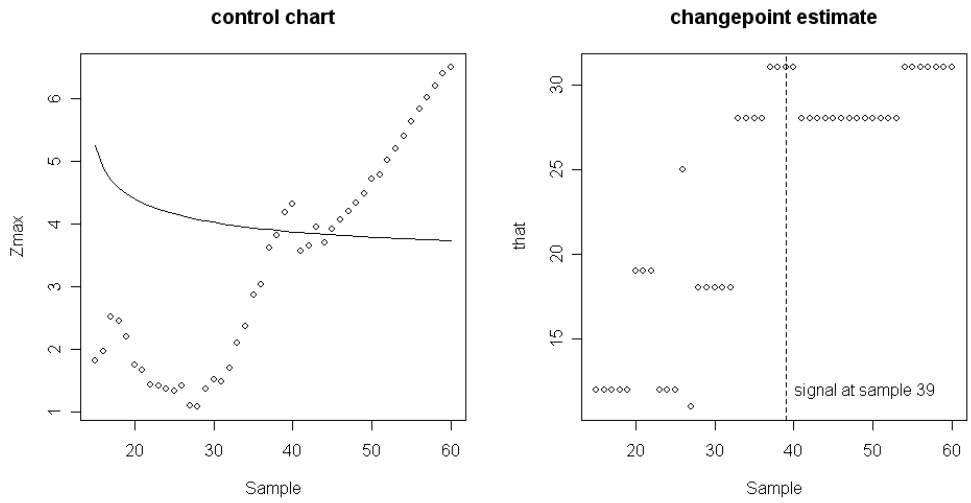

Figure 4.2: Normal changepoint model on log transformed SiO2 data: $Z_{\max,n}$ and $h_{n,0.002}$

be appropriate. We explored these issues by simulating three settings: shift occurs at 15th observation, 50th observation or 500th observation, which means $\tau = 14$, $\tau = 49$ or $\tau = 499$. For all, we simulated 200,000 sequences. Testing started at the 15th process reading, and used control limits giving an in-control ARL of 500.

The process readings simulated were $N(0, 1)$ up to time τ, and $N(\delta, 1)$ after time τ. Table 4.3 shows the ARL of both parametric and nonparametric change-point chart. Figure 4.3 represents the numbers in a plot which is easier for comparison. The figure shows some expected features. Having 49 rather than 14 initial in-control readings speeds the detection of a shift. The benefit is greatest when the shift is small. It is also interesting that having 499 initial in-control readings still gives a considerable improvement over having 49 initial readings when the shift is small.

A more interesting difference is between parametric and nonparametric. It is natural to expect that nonparametric change-point model could beat the parametric for a non-normal, especially skewed, underlying distribution. When the true distribu-

tion is normal however, one would expect the parametric method to outperform the nonparametric across the board, though substantially so only for large shifts. Surprisingly, however, the nonparametric formulation offers you more than that. For $\tau = 14$, $\tau = 49$ and $\tau = 499$, the nonparametric model has a shorter ARL when shift is within 2.1, 1.5 and 1 standard deviations respectively. This means that for small to moderate shift, the nonparametric change-point model outperforms the parametric even when the underlying distribution is normal. For large shift, intuition is right and the parametric approach is better.

It is quite a surprise to see that the nonparametric beats the parametric for the small to moderate shift for the normal underlying distribution. The reason and the remedy will be explored in Chapter 6.

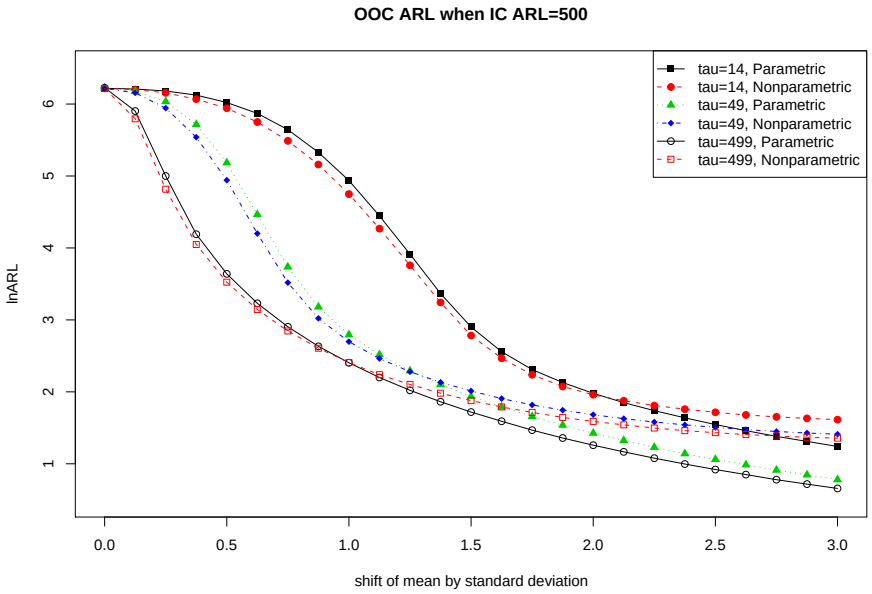

Figure 4.3: The ARL's of the parametric and nonparametric change-point procedure when a mean shift of size δ is introduced from 15th, 50th and 500 observation.

	α					
n	0.02	0.01	0.005	0.002	0.001	0.0005
15	2.700	2.848	2.947	3.069	3.181	3.229
16	2.615	2.767	2.91	3.047	3.142	3.244
17	2.535	2.718	2.862	3.043	3.163	3.247
18	2.535	2.694	2.86	3.034	3.183	3.277
19	2.500	2.695	2.869	3.054	3.186	3.296
20	2.488	2.699	2.851	3.059	3.203	3.311
22	2.468	2.692	2.862	3.082	3.228	3.355
24	2.469	2.676	2.870	3.096	3.249	3.389
26	2.452	2.686	2.875	3.108	3.269	3.415
28	2.455	2.686	2.883	3.121	3.283	3.437
30	2.453	2.684	2.879	3.13	3.297	3.453
35	2.452	2.687	2.894	3.149	3.324	3.487
40	2.447	2.689	2.900	3.162	3.342	3.511
45	2.453	2.690	2.906	3.171	3.356	3.529
50	2.451	2.691	2.908	3.178	3.365	3.542
60	2.452	2.694	2.914	3.188	3.379	3.560
70	2.452	2.694	2.917	3.194	3.388	3.570
80	2.453	2.696	2.918	3.199	3.394	3.579
90	2.452	2.696	2.920	3.200	3.399	3.584
100	2.453	2.697	2.922	3.203	3.402	3.591
125		2.698	2.923	3.206	3.409	3.599
150		2.697	2.924	3.209	3.411	3.603
175		2.698	2.924	3.210	3.414	3.604
200		2.699	2.926	3.210	3.415	3.610
250		2.700	2.927	3.212	3.416	3.610
300		2.704	2.926	3.215	3.420	3.616
500			2.927	3.213	3.417	3.612
1000			2.927	3.214	3.418	3.612

Table 4.1: Cutoffs $h_{\alpha,n}$ for sample size n and Hazard Rate α starting at Sample 15

n	SiO2	n	SiO2	n	SiO2	n	SiO2
1	0.27	16	0.15	31	0.23	46	0.58
2	0.09	17	0.07	32	0.51	47	0.57
3	1.55	18	0.19	33	0.73	48	0.54
4	0.18	19	0.27	34	0.52	49	0.65
5	0.17	20	0.77	35	0.88	50	1.04
6	0.18	21	0.34	36	0.49	51	0.48
7	0.44	22	0.24	37	1.28	52	1.16
8	0.36	23	0.10	38	0.59	53	0.88
9	0.27	24	0.26	39	0.81	54	1.04
10	0.29	25	0.25	40	0.55	55	1.68
11	0.29	26	0.62	41	0.12	56	1.07
12	0.23	27	0.17	42	0.44	57	2.72
13	0.10	28	0.27	43	0.98	58	1.06
14	0.26	29	0.56	44	0.21	59	1.24
15	0.07	30	0.41	45	0.71	60	0.65

Table 4.2: Example of silica concentration

$\tau = 14$			$\tau = 49$			$\tau = 499$		
shift(δ)	HQK	NPC	shift(δ)	HQK	NPC	shift(δ)	HQK	NPC
0	502.16	501.08	0	504.18	501.59	0	506.04	501.34
0.125	497.09	493.52	0.125	480.66	471.60	0.125	366.10	328.55
0.25	483.66	471.51	0.25	417.86	381.19	0.25	148.38	123.52
0.375	456.91	431.80	0.375	303.75	254.45	0.375	66.09	57.30
0.5	412.01	380.67	0.5	178.50	140.06	0.5	38.09	33.93
0.625	354.48	314.36	0.625	86.92	66.71	0.625	25.30	23.20
0.75	282.55	241.91	0.75	41.92	33.72	0.75	18.26	17.22
0.875	206.27	174.13	0.875	24.05	20.53	0.875	13.91	13.55
1	139.35	115.43	1	16.34	14.84	1	11.07	11.11
1.125	85.24	71.46	1.125	12.39	11.75	1.125	9.02	9.38
1.25	50.09	42.89	1.25	9.90	9.78	1.25	7.55	8.19
1.375	28.98	25.61	1.375	8.20	8.44	1.375	6.44	7.25
1.5	18.20	16.17	1.5	6.93	7.48	1.5	5.58	6.54
1.625	12.91	11.79	1.625	5.99	6.73	1.625	4.91	5.99
1.75	10.07	9.37	1.75	5.25	6.16	1.75	4.34	5.55
1.875	8.41	7.96	1.875	4.65	5.73	1.875	3.89	5.18
2	7.24	7.10	2	4.16	5.38	2	3.52	4.90
2.125	6.36	6.53	2.125	3.76	5.09	2.125	3.21	4.67
2.25	5.69	6.11	2.25	3.42	4.86	2.25	2.94	4.47
2.375	5.15	5.81	2.375	3.13	4.67	2.375	2.71	4.31
2.5	4.69	5.56	2.5	2.89	4.51	2.5	2.51	4.19
2.625	4.31	5.36	2.625	2.68	4.37	2.625	2.34	4.08
2.75	3.98	5.22	2.75	2.49	4.26	2.75	2.18	4.00
2.875	3.71	5.11	2.875	2.33	4.17	2.875	2.05	3.93
3	3.46	5.02	3	2.18	4.10	3	1.93	3.88

Table 4.3: Comparison of ARL for parametric and nonparametric change-point models; IC ARL=500; start testing at Sample 15; a mean shift of size δ is introduced from 15th, 50th and 500 observation.

Chapter 5

Asymptotic Results in Phase II Analysis

5.1 Parametric Change-Point Model

In HQK, it is observed that the Control limit appears to become stable when n is large enough ($n > 400$ conservatively). Intuitively, it makes sense in that, when you have at least one sample being infinitely large, the asymptotic distribution of two-sample t-statistic should not change when you add one more observation into the larger sample. However, there were no theoretically based asymptotic results available up to this point.

Actually, for the case when post-change parameter is unknown, which corresponds to the generalized likelihood ratio rule as (3.1.5), the asymptotic properties remained a standing problem for a long time until Siegmund and Venkatraman (1995) provided the following approximation.

When the in-control distribution mean is μ_o and the variance keeps to be σ^2, an alternative form of (3.1.5) is

$$N_G = \inf\{n : \max_{0\le k\le n-1} \frac{|\bar{X}_{k,n} - \mu_o|}{\sigma\sqrt{\frac{1}{n-k}}} \ge b\}, \tag{5.1.1}$$

where $\bar{X}_{k,n}$ is defined in (2.1.4) and is the mean of last $n-k$ observations.
When there is no change, i.e. the process is always in control, then as $b \to \infty$, N_G is asymptotically exponentially distributed with expectation

$$E(N_G) \sim \frac{(2\pi)^{1/2}\exp(b^2/2)}{b\int_0^\infty x\nu^2(x)dx}, \tag{5.1.2}$$

or numerically more accurate but asymptotically equivalent form:

$$E(N_G) \sim \frac{(2\pi)^{1/2}\exp(b^2/2)}{b\int_0^b x\nu^2(x)dx}. \tag{5.1.3}$$

where $\nu(x)$ is defined in (2.1.14).
Notice the in-control ARL $E(N_G)$, i.e. $1/\alpha$, is a function only depends on b. So for any given α, the corresponding control limit b could be found by solving the above equation.

In section 3.3, we have seen a connection between generalized CUSUM and HQK. With a trivial reformatting, (3.2.2) could be represented as

$$N_{HQK} = \inf\{n : \max_{1\le k\le n-1} \frac{|\bar{X}_{k,n} - \bar{X}_{0,k}|}{\hat{\sigma}\sqrt{\frac{1}{n-k} + \frac{1}{k}}} \ge h_{\alpha,n}\}, \tag{5.1.4}$$

where $\bar{X}_{0,k}$ and $\bar{X}_{k,n}$ are defined in (2.1.3) and (2.1.4), which are the mean of the first k and the last $n-k$ observations respectively.

Provided that $n \to \infty$ and b is relatively high, the left segment mean $\bar{X}_{0,k} \to \mu_o$, $1/k \to 0$ and $\hat{\sigma} \to \sigma$. To the extent that $\frac{|\bar{X}_{0,k} - \bar{X}_{k,n}|}{\hat{\sigma}\sqrt{\frac{1}{n-k}+\frac{1}{k}}}$ is distributed like $\frac{|\bar{X}_{k,n} - \mu_o|}{\sigma\sqrt{\frac{1}{n-k}}}$, the cutpoint $h_{\alpha,n}$ should be approximated by the constant cutpoint b. However, the formal proof is not particularly trivial.

In table 5.1, we compared the values of b in (5.1.2) or (5.1.3) with the last row of $\tilde{h}_{\alpha,n}$'s in parametric table. Since when to start testing has little effect on the limit of $\tilde{h}_{\alpha,n}$, without loss of generality, we use the ones where we start testing from 21st observation. Technically, we implemented the integration by adaptive quadrature using "integrate" function of R and implemented the root-finding by bisection search algorithm using "bisearch" function in "extRemes" module of R. The results are also displayed in Figure 5.1, with the solid lines showing the $\tilde{h}_{\alpha,n}$'s and the dashed horizontal lines showing the asymptotic b values for the corresponding ARL, both correspond to $\alpha = 0.001, 0.002, 0.005, 0.01, 0.02$ top down. From the plot, the values of b in (5.1.2) or (5.1.3) matched the limit of $\tilde{h}_{\alpha,n}$'s very well, especially for small α, i.e. large b values. This observation then provides both a rationale for HQK's observation that the $\tilde{h}_{\alpha,n}$ values appeared to be stabilizing, and an explicit formula for the value to which they stabilize.

5.2 Nonparametric Change-Point Model

The control limits of nonparametric change-point model are displayed in Figure 5.2, for $\alpha = 0.001, 0.002, 0.005, 0.01, 0.02$. Unlike the monotonely decreasing control limit in parametric model, the nonparametric has a v-shaped control limit. The values decrease at first, then increase slowly until stabilized. This trend shows up from $\alpha = 0.005$ and becomes more evident for smaller false signal rate α.

Although the control limit of nonparametric model seems to become stable after a while, just like the parametric case, the value of the asymptotic limit is unknown

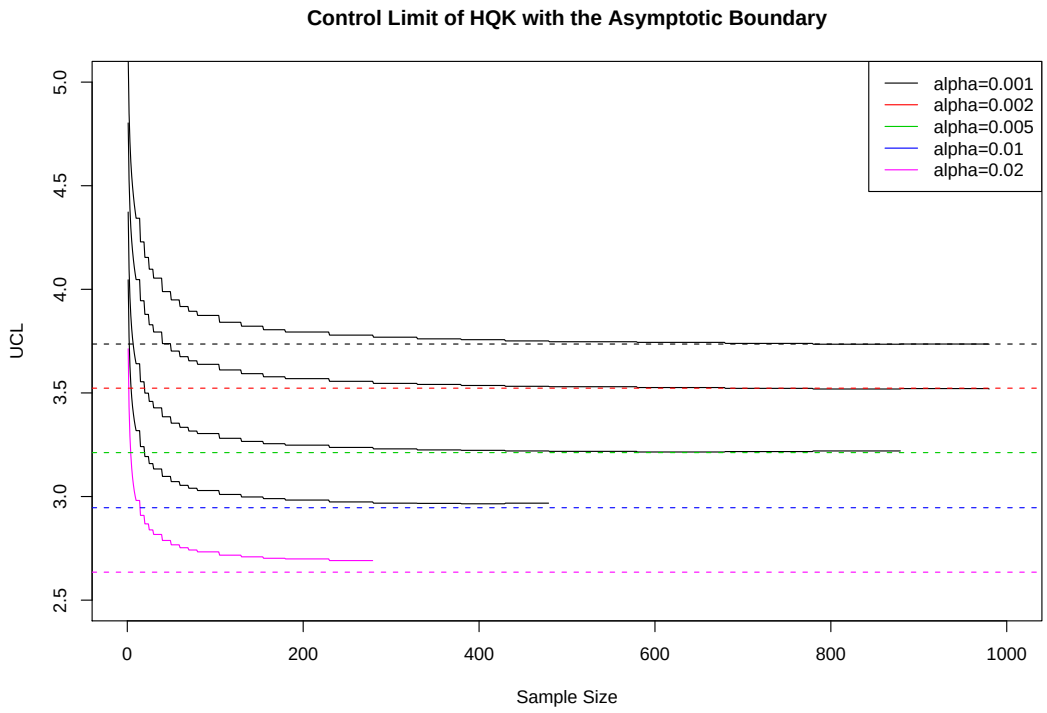

Figure 5.1: The b values and the Control limit of HQK for $\alpha = 0.001, 0.002$, 0.005, 0.01, 0.02.

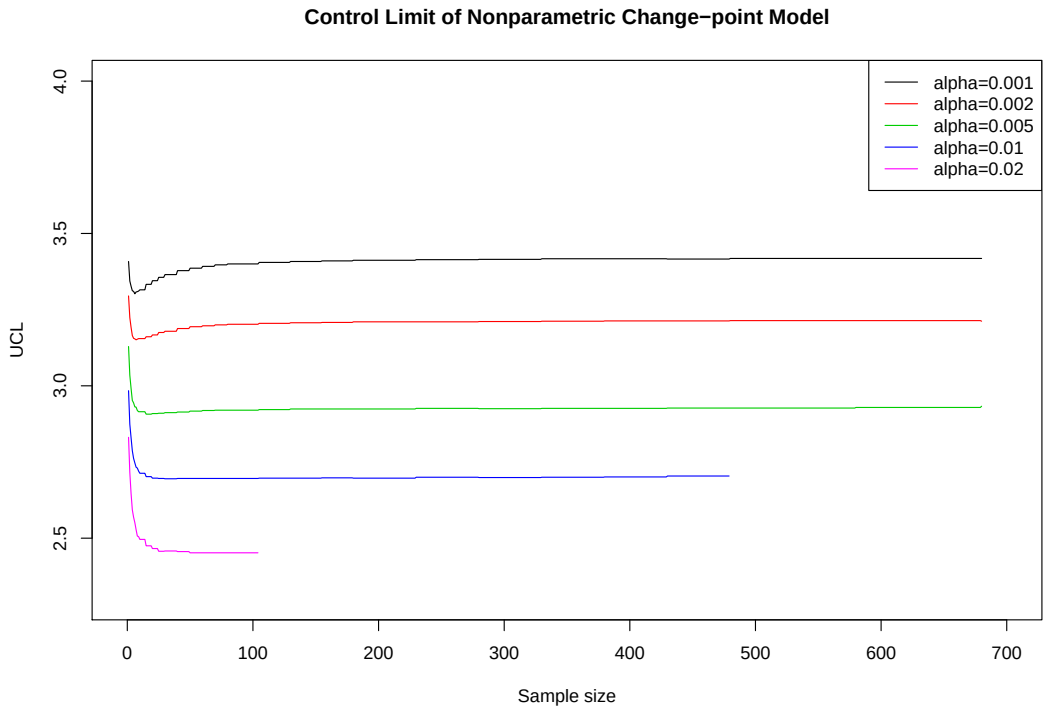

Figure 5.2: The Control limit of NPC for $\alpha = 0.001, 0.002, 0.005, 0.01, 0.02$.

at this point. Table 5.1 compared the last row of $h_{\alpha,n}$'s in nonparametric table with the last row of $\tilde{h}_{\alpha,n}$'s in parametric. We can see that when α gets smaller, $h_{\alpha,n}$ gets closer to some asymptotic value, it is clear that this asymptotic value is different from the value for the corresponding parametric statistic. This contrasts with the Phase I situation in which the parametric and nonparametric tests have the same asymptotic fractiles. Note that the asymptotic value for the nonparametric statistic is considerably smaller than the parametric. For example the asymptotic value for an ARL of 200 with the parametric test gives an ARL in excess of 500 for the nonparametric.

α	$h_{\alpha,n}$	$\tilde{h}_{\alpha,n}$	b
0.02	2.452	2.691	2.635
0.01	2.704	2.968	2.946
0.005	2.933	3.220	3.212
0.002	3.212	3.519	3.523
0.001	3.418	3.735	3.736

Table 5.1: The comparison of b values, the limit of $\tilde{h}_{\alpha,n}$'s and the limit of $h_{\alpha,n}$'s

Chapter 6

Quarantined change-point model and Window-limited approach

6.1 Window-limited Approach

In the change-point methodology as described, at observation n one tests the entire sequence of data so far for a change point. This entails explicitly or implicitly storing all n values, and performing $n-1$ left-right comparisons. In the past, both the storage and the computational aspects of this became onerous for large n values – say in the thousands. To avoid this, Willsky and Jones (1976) proposed the window-limited approach. In this, the traditional GLR rule is modified to the window-limited GLR rule:

$$N_w = \inf\{n : \max_{n-M \le k \le n-\tilde{M}} \sup_{\theta} \sum_{i=k+1}^{n} \log \frac{f_\theta(X_i)}{f_{\theta_o}(X_i)} \ge c_\gamma\}. \tag{6.1.1}$$

In the window-limited approach, $\tilde{M}$ is usually set to be 1 so there is no quarantine at the right edge. However, an appropriate M will be used to exclude a larger segment at the left edge, leaving only M observations working in the formula. The choice of M

is not trivial at all. Lai (2001) provided a discussion about it. This method eliminates the ability to signal a change occurring in the first $n-M-1$ time points. However, it is reasonable since no signal up to this point implies the process likely stayed in-control for the remote past, otherwise we should have a signal. Since the window-limited approach throws away a large part of data, it loses part of the information. However, considering the trade-off between the computational burden and the loss of efficiency, the sacrifice may be worthwhile, especially at 30 years ago.

6.2 Quarantined Change-Point Model

We noted that the parametric and nonparametric statistics $Z_{k,n}$ and $T_{k,n}$ differ primarily in that when k or $n-k$ is small, $Z_{k,n}$ follows a normal distribution but $T_{k,n}$ does not. This suggests that differences in behavior of the change-point test statistics must be driven largely by the impact of considering very short segments. That this impact can be considerable was shown quite starkly by the comparison of the asymptotic control limits, where the parametric limits were substantially larger than the nonparametric. As another indication, recall that in the Phase I setting, restricting the search range of k to $cn < k < (1-c)n$ for any $0 < c < 1/2$ was sufficient to change the large-n distribution of $Z_{\max,n}$ from an unbounded extreme value to a bounded asymptotic distribution. These observations motivate considering a corresponding modification in Phase II testing to exclude testing of very short segments on the left or right.

This method is different from the traditional window-limited approach in various ways. First, rather than reducing the computational burden, this method is targeted to improve the performance of change-point model. Second, comparing with dropping an increasing size of segment on the left, we only drop small and fixed number of test statistics. Actually, this method focuses more at the segment on the right, since the

signal on the left side will become rarer and rarer given that there is no signal on all previous tests. So quarantine on the left is not expected to make as much difference as quarantine on the right. Due to these essential differences, we would like to name it quarantined change-point model to distinguish with the traditional window-limited approach.

Mathematically, we propose to use

$$N_{qHQK} = \inf\{n : \max_{c<k<n-c} \tilde{T}_{k,n} \geq \tilde{h}_{\alpha,n,c}\}, \tag{6.2.1}$$

where c is the number of $\tilde{T}_{k,n}$ quarantined at the left and at the right. $\tilde{T}_{k,n}$ is defined in (2.1.8). And $\tilde{h}_{\alpha,n,c}$ should be set such that this conditional probability of a false alarm at any n is a constant α.

We also do the same quarantine action for nonparametric change-point model:

$$N_{qNPC} = \inf\{n : \max_{c<k<n-c} T_{k,n} \geq h_{\alpha,n,c}\}, \tag{6.2.2}$$

where c is the number of $T_{k,n}$ quarantined at the left and at the right. $T_{k,n}$ is defined in (2.2.6). And similarly, $h_{\alpha,n,c}$ should be set such that this conditional probability of a false alarm at any n is a constant α.

It should be noticed that unlike the window-limit approach, we still use all the data available to calculate the test statistics. The only thing dropped is the ability of splitting at a small segment for either end of the sequence. Actually, we are still making full use of the existing data.

We investigated c values of 3, 6 and 9, in addition to the no-quarantine $c = 0$. For each setting of these quarantined versions of the test statistics, forty million sequences of length 1000 were simulated to find cutpoints up to $n = 1000$. The results are displayed in the appendix. In the parametric case, these cutpoints are presented in Table $1 \sim 13$ of appendix, start testing from the first possible test observation, 10, 15 and 21 if possible, for each c value in $0, 3, 6, 9$. In the nonparametric case, these

cutpoints are presented in Table 14 $\sim$ 20 of appendix, start testing from 15 or the first possible test observation, and 21, for each c value in $0, 3, 6, 9$. The control limit for nonparametric model without quarantine has been shown in Table 4.1, so it is not included here. The columns corresponding to in-control average run lengths of 20, 50, 100, 200, 500 and 1000 respectively for parametric, and to 50, 100, 200, 500, 1000 and 2000 respectively for nonparametric. Notice that when $c = 9$, you will be not able to start test until you have 19 initial observations, no matter how many warm-up cases you originally designed. In order to make it comparable for different c, for all the three tables, we list the control limits with 20 warm-up data, i.e. the method starts actual monitoring from the 21st process reading. Again, not all n values are listed explicitly; you can carry entries forward.

One striking feature of these tables is the degree to which quarantining lowers the control limits in parametric model. This suggests that modest shifts should be detected faster, as the test statistic has a lower hurdle to cross. For example, the parametric cutoffs for $n = 100$ with 20 initial warm-up observations are shown is Table 6.1.

			α					
c	initial	n	0.05	0.02	0.01	0.005	0.002	0.001
0	20	100	2.301	2.733	3.029	3.304	3.638	3.874
3	20	100	1.969	2.463	2.795	3.102	3.470	3.723
6	20	100	1.766	2.283	2.639	2.964	3.349	3.614
9	20	100	1.619	2.150	2.518	2.854	3.253	3.529

Table 6.1: Cutoffs $\tilde{h}_{\alpha,100,c}$ for Hazard Rate α starting at Sample 21, Parametric.

From the table, the cutoffs with quarantining are more or less like the cutoffs

without quarantining shifting one column to the right. For instance, $\tilde{h}_{0.01,100,0} = 3.029$ (IC ARL=100) is close to $\tilde{h}_{0.01,100,6} = 2.964$ (IC ARL=200). In parametric model, the in-control sequences have strong tendency to split with a short segment with high extreme values, so chopping off those split points will considerably lowers the cutoffs. However, the out-of-control sequences with moderate shift need to wait and accumulate information before being able to signal, so it is inclined to split with a longer segment. In this sense, the quarantining will have much less effect on the out-of-control distributions of the maximum statistic, so the approximately even $\tilde{h}_{0.01,100,0}$ and $\tilde{h}_{0.01,100,6}$ imply the approximately comparable OOC ARLs. Along with the doubled IC ARLs, that is how quaranting brings a great benefit for moderate shift.

In contrast, for nonparametric table 6.2, the quaranting has very little effect on the cutoffs. This make sense in that, the nonparametric model itself has much less inclination to split near the end of the sequence. So it will not benefit much from quarantining. All the results will be confirmed in the next section.

			α					
c	initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
0	20	100	2.452	2.696	2.920	3.202	3.400	3.587
3	20	100	2.414	2.684	2.921	3.203	3.402	3.585
6	20	100	2.265	2.579	2.856	3.176	3.390	3.582
9	20	100	2.140	2.474	2.773	3.115	3.344	3.553

Table 6.2: Cutoffs $h_{\alpha,100,c}$ for Hazard Rate α starting at Sample 21, Nonparametric.

6.3 Evaluation of Performance

In this section, we will investigate the effect of quarantine on both nonparametric and parametric change-point models. For each model, we consider four cases: $c = 0, c = 3, c = 6, c = 9$, where $c = 0$ corresponds to the model without quarantine and $c = 3/6/9$ corresponds to the model with $3/6/9$ $T_{k,n}$'s or $\tilde{T}_{k,n}$'s dropped on each side. When $c = 9$, we will need at least 20 observations in order to perform the test. So for comparison purpose, we start testing from 21st observation for all the cases.

Literally, in the simulation, for both parametric and nonparametric, we consider eight settings: shift occurs at 21st observation or 50th observation, which means $\tau = 20$ or $\tau = 49$; each with the variation of $c = 0, 3, 6$ or 9. In all the simulation, testing started at the 21st process reading, and the control limits are set for an in-control ARL of 500. For all, we simulated 200,000 sequences.

The process readings simulated were $N(0, 1)$ up to time τ, and $N(\delta, 1)$ after time τ. Figure 6.1 presents the resulting ARL's in HQK for δ values in the range of $[0, 3]$, while figure 6.2 presents the same thing in nonparametric change-point model.

From Figure 6.1, we see that in the parametric change-point model, quarantining a few observations gives you a big improvement for small to moderate shifts (up to about $1.5 \sim 2$ standard deviation), but hurts for large shift. The improvement is especially noticeable from no quarantine to $c = 3$. The reason is that we eliminate the possibility of splitting at the two ends of the sequence. So when a large shift happens, the parametric model without quarantine can react quickly, while the quarantined model has to wait for a while before being able to signal. However, the extreme values in the non-quarantined model result in a considerably higher control limits which is harder to reach when the shift is just moderate.

This phenomenon reminds us of the performance comparison between parametric and nonparametric in section 4.4. We have seen that for small to moderate shift, the

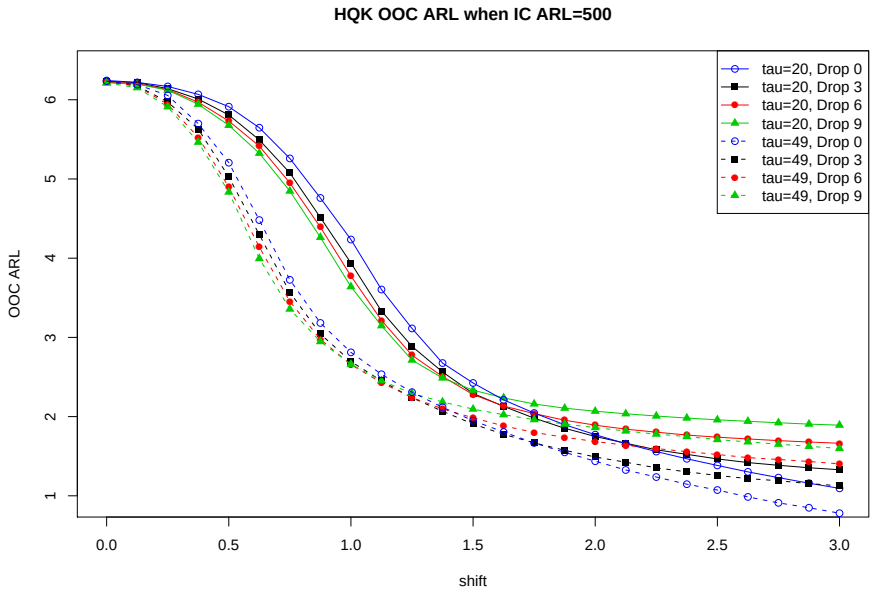

Figure 6.1: The ARL's of the parametric changepoint procedure when a mean shift of size δ is introduced from 21st, and 50th observation.

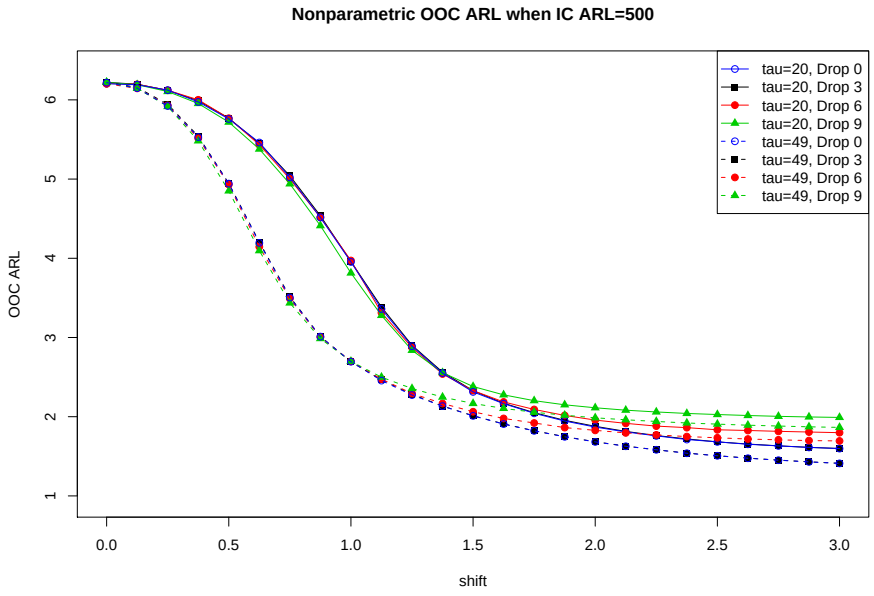

Figure 6.2: The ARL's of the nonparametric changepoint procedure when a mean shift of size δ is introduced from 21st and 50th observation.

nonparametric change-point model beats the parametric even when the data follows normal distribution. This could be explained in the same manner. Compared with nonparametric model, HQK has a strong tendency to split off one or two observations at each end of the sequence. And those $T_{k,n}$'s at the end of sequence tend to be more extreme than the corresponding $\tilde{T}_{k,n}$'s. So when a large shift happens, the parametric model can react faster than nonparametric. However, the higher control limits brings a slower signal for small shift.

From Figure 6.2, it is not very surprising to see the quarantining has little effect in nonparametric model, since the control limits are only slightly changed by quarantining.

Again, having more initial in-control readings (49 instead of 20) speeds the detection of a shift and the benefit is greatest when the shift is small.

Chapter 7

Future Work

7.1 Robustness of Parametric Change-Point Model.

HQK is developed based on the assumption the underlying distribution in normal. For non-normal data, HQK is not appropriate. The traditional parametric change-point model is not robust with respect to normal distribution assumption, since it tends to split off with a short segment on one side, where the normality is clearly violated, and the IC ARL will be quite altered from the expected value. However, it will be interesting to investigate the robustness of the quarantined parametric change-point model. For quarantined model, the two segments on either side of the splitting point are forced to be larger than c, where the central limit theory starts to take action on both sides. Then the sample mean of the two segments will be more like a normal shaped distribution even if the true distribution of the single observation is not normal.

7.2 Asymptotic control limit of Change-Point Model

Although we have found the b value matched very well with the asymptotic limit of HQK cutoff $\tilde{h}_{\alpha,n}$, the formal proof is not provided. Meanwhile, it it shown in chapter 5 that the asymptotic control limit of nonparametric cutoff is distinctly different from that of HQK and the b value, but what is the asymptotic value of $h_{\alpha,n}$ is still a mystery. As far as we know, there is no theoretical foundation to support a specific value up to this point.

7.3 Asymptotics of $\hat{\tau}$ for small shifts under H_a

There has been some work on the estimate of the change point in the Phase I setting. The model used is somewhat unrealistic, in that τ is considered fixed, while the data series grow on both the left and the right. Under the H_a case where there is a real persistent small shift occurs at τ, Hinkley (1970) showed in the Phase I setting,

$$a_n(\hat{\tau}_Z - \tau) \to G.$$

G is a known nondegenerate distribution, and a_n is a known sequence depending on n.

For nonparametric case, when there is a small shift, and it diminishes at a rate no faster than $n^{-1/2}$, Ferger(1994), also in the Phase I setting, claimed that

$$b_n(\hat{\tau}_T - \tau) \to H.$$

Again, H is a nondegenerate distribution. And b_n is a sequence depending on n.

For a persistent small shift, the behavior is expected to be similar between those two. The two proofs are based on the construction of two defective random walks, one from change-point τ to the right, the other from τ to the left. Both of them

can be represented as sum of some independent, identically distributed increment process. So far, we have found out the random walks for $Z_{k,n}$ and $T_{k,n}$ have very similar structure. However, further work needs to be done before we come up with some meaningful results.

When the similar convergence structure is confirmed, the outcome is very likely to be that G and H are the same, given all the identical properties we have mentioned. However, the sequences a_n and b_n will likely differ. It will be interesting to work on the explicit expression for a_n and b_n, leading to the efficiency comparison between $Z_{k,n}$ and $T_{k,n}$.

To date however, no theoretical work appears to have been carried out on the estimated change-point of either the parametric or the nonparametric procedures in Phase II, making this a completely blank slate.

Chapter 8

Summary

The Phase I non-parametric change-point model based on weighted U statistics has a very similar behavior to the one based on normal Z score. They have the same correlation structure and the same asymptotics when the real distribution is normal. This suggests those two are closely parallel with each other. There is extensive literature on the normal change-point model, and its theoretical foundation has been explored in detail . The Phase I behavior of the nonparametric counterpart is less well studied, and the relationship between them is even less well studied. So with these parallels, the normal case provides us a valuable guideline to fill in the corresponding gaps in nonparametric models, and it is a useful tool for us to understand the unknown behavior of nonparametric models.

The introduction of the parametric change-point model into Phase II SQC has brought a whole new perspective to the problem of unknown parameters, and our nonparametric change-point model takes this a step further. In statistical theory, a simple two-sample Mann-Whitney test has a relative efficiency as high as 96% when the distribution is normal, while it is a great improvement when the true distribution is not. Our results show something even better - that for moderate shifts there appears

to be no loss of performance in using a nonparametric rather than a parametric approach. Moreover, it requires minimal assumptions and knowledge of baseline data to detect a change in an ongoing process, and can give the greatest freedom from worry about the underlying distribution. This method extends the applicability of previous proposals for Phase II change-point model.

Another interesting finding is the quarantined parametric change-point model. By eliminate the ability of splitting at a small segment for either end of the sequence, parametric will signal faster for small to moderate shift, but slower for large shift. For nonparametric model, quarantining doesn't make much difference.

Bibliography

[1] Amin, R. W., Reynolds, M. R., and Bakir, S. T. (1995) Nonparametric Quality Control Charts Based on the Sign Statistic. *Communications in Statistics-Theory and Methods*, **24**, 1597-1623.

[2] Amin, R. and Searchy, A. J. (1991) A Nonparametric Exponentially Weighted Moving Average Control Scheme. *Communications in Statistics-Simulation and Computation*, **20**, 1049-1072.

[3] Barnard, G. A. (1959) Control charts and stochastic processes (with discussion). *J. Roy. Statst. Soc. Ser. B*, **21**, 239-271.

[4] Bakir, S. T. and Reynolds, M. R. (1979) A Nonparametric Procedure for Process Control Based on Within-Group Ranking. *Technometrics*, **21**, 175-183.

[5] Chakraborti, S., Van Der Lann, P. and Bakir, S.T. (2001) Nonparametric Control Charts: An Overview and Some Results. *Journal of Quality Technology*, **Vol. 33, No. 3**, 304-315.

[6] Csörgő, M. and Horváth, L.(1997) *Limited Theorems in Change-Point Analysis*. Chichester: Wiley.

[7] Ferger, D. On the power of nonparametric changepoint-tests. *Metrika*, **41**, 277-292.

[8] Hackl, P. and Ledolter, J. (1991) A Control Chart Based on Ranks. *Journal of Quality Technology*, **23**, 117-124.

[9] Hawkins, D.M. (1977) Testing a sequence of Observations for a shift in Location. *Journal of the American Statistical Association*, **Vol. 72, No. 357**, 180-186.

[10] Hawkins, D.M., Qiu, P., Kang, C.W. (2002) The Changepoint Model for Statistical Process Control. *Journal of Quality Technology*, **35**, 355-365.

[11] Hinkley, D.V. (1970) Inference about the change-point in a Sequence of Random Variables. *Biometrika*, **Vol.57, No.1**, 1-17.

[12] Irvine, J.M. (1982) *Ph.D. dissertation: Changes in Regime in Regression models.* Yale University.

[13] James, B., James, K. L. and Siegmund, D.(1987) Tests for a Change-Point. *Biometrika*, **76**, 71-83.

[14] Janacek, G. J. and Meikle, S. E. (1997) Control Charts Based on Medians. *The Statistician*, **46**, 19-31.

[15] Jesen, W. A. Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006) Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, **38**, 349-364.

[16] Jones, L. A., Champ, C. W. and Rigdon, S. E. (2001) The performance of Exponentially Weighted Moving Average Charts with Estimated Parameters. *Technometrics*, **43**, 156-167.

[17] Jones, L. A., Champ, C. W. and Rigdon, S. E. (2004) The Run Length Distribution of the CUSUM with Estimated Parameters. *Journal of Quality Technology*, **36**, 95-108.

[18] Lai, T.L. (2001) Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, **11**, 303-408.

[19] McDonald, D. (1990) A CUSUM Procedure Based on Sequential Ranks.. *Naval Research Logistics*, **37**, 627-646.

[20] Page, E. S. (1954) Coninuous Inspection Schemes. *Biometrika*, **41**, 100-114.

[21] Pettitt, A.N. (1979) A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, **Vol. 28, No. 2**, 126-135.

[22] Roberts, S.W. (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, **8**, 411-430.

[23] Schechtman, E. and Wolfe, D.A. (1981) Distribution-free tests for the change-point problem. *Technical Report, The Ohio State University*, , .

[24] Schechtman, E. (1982) A Nonparametric Test for Detecting Changes in Location. *Communications in Statistics - Theory and Methods*, **11(13)**, 1475-1482.

[25] Schechtman, E. and Wolfe, D.A. (1984) Nonparametric Statistical Procedures for the Changepoint Problem. *Journal of Statistical Planning and Inference*, **9**, 389-396.

[26] Shewhart, W. A. The application of statistics as an aid in maintaining quality of a manufactured product. *Journal of the American Statistical Association*, **20**, 546-8.

[27] Siegmund, D. and Venkatraman, E. S. (1995) Using the Generalized Likelihood Ratio Statistic for Sequential Detection of a Change-point. *The Annals of Statistics*, **Vol. 23, No. 1**, 255-271.

[28] Willemain, T. R. and Runger, G. C.(1996) Designing Control Charts Using an Empirical Reference Distribution. *Journal of Quality Technology*, **28**, 31-38.

[29] Willsky, A. S. and Jones, H. G. (1976) A generalized likelihood ratio approach to detection and estimation of jumps in linear systems. *IEEE Trans. Automat. Contr.*, **21**, 108-112.

[30] Woodall, W. H. (2000) Controversies and Contradictions in Statisitical Process Control (with discussion). *Journal of Quality Technology*, **32**, 341-378.

[31] Woodall, W. H. and Montgomery, D. C. (1999) Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology*, **31**, 376-386.

[32] Zhou, C., Zou, C., Zhang, Y., Wang, Z. (2007) Nonparametric Control Chart Based on Changepoint Model. *Statistical Papers*, , .

Appendices: Tables of Control Limits for Quarantined Change-Point Model

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
2	3	38.140	95.368	190.740	381.429	953.534	1907.334
2	4	7.320	11.826	16.927	24.127	38.460	54.735
2	5	4.876	6.910	8.907	11.429	15.763	20.040
2	6	4.058	5.399	6.618	8.057	10.368	12.508
2	7	3.620	4.695	5.595	6.608	8.157	9.527
2	8	3.344	4.275	5.023	5.829	7.010	8.021
2	9	3.158	3.992	4.647	5.337	6.319	7.124
2	10	3.024	3.790	4.383	4.995	5.850	6.537
2	11	2.923	3.640	4.187	4.745	5.510	6.118
2	12	2.846	3.524	4.036	4.554	5.256	5.807
2	13	2.783	3.433	3.918	4.405	5.056	5.564
2	14	2.733	3.358	3.820	4.282	4.895	5.366
2	15	2.690	3.296	3.742	4.183	4.765	5.211
2	16	2.654	3.244	3.676	4.099	4.657	5.078
2	17	2.624	3.200	3.619	4.030	4.566	4.972
2	18	2.597	3.162	3.570	3.969	4.486	4.879
2	19	2.574	3.129	3.528	3.917	4.418	4.795
2	20	2.554	3.100	3.492	3.871	4.359	4.723
2	21	2.536	3.073	3.459	3.830	4.309	4.661
2	22	2.520	3.051	3.430	3.795	4.261	4.607
2	23	2.506	3.030	3.403	3.762	4.221	4.561
2	24	2.493	3.011	3.380	3.733	4.183	4.514
2	25	2.481	2.994	3.358	3.707	4.150	4.475
2	26	2.470	2.979	3.339	3.683	4.119	4.440
2	27	2.460	2.965	3.322	3.661	4.091	4.408
2	28	2.451	2.952	3.305	3.640	4.065	4.374
2	29	2.443	2.940	3.290	3.622	4.041	4.348
2	30	2.436	2.930	3.277	3.606	4.020	4.321
2	35	2.405	2.884	3.221	3.537	3.933	4.219
2	40	2.383	2.854	3.182	3.489	3.872	4.148
2	45	2.365	2.830	3.151	3.453	3.824	4.093
2	50	2.353	2.811	3.128	3.424	3.791	4.051
2	60	2.334	2.784	3.094	3.382	3.737	3.988
2	70	2.322	2.765	3.070	3.353	3.701	3.948
2	80	2.310	2.751	3.053	3.333	3.675	3.917
2	90	2.304	2.740	3.040	3.316	3.655	3.893
2	100	2.301	2.732	3.028	3.304	3.638	3.874
2	125		2.715	3.009	3.281	3.610	3.841
2	150		2.708	2.998	3.266	3.592	3.822
2	175		2.701	2.989	3.254	3.578	3.805
2	200		2.702	2.983	3.248	3.569	3.793
2	250		2.692	2.973	3.237	3.556	3.779
2	300			2.966	3.229	3.546	3.769
2	350			2.967	3.225	3.541	3.761
2	400			2.966	3.223	3.536	3.757
2	450			2.970	3.220	3.532	3.751
2	500				3.217	3.530	3.748
2	600				3.216	3.526	3.744
2	700				3.216	3.522	3.739
2	800					3.520	3.736
2	900					3.521	3.735
2	1000					3.518	3.734

Table 1: Cutoffs $\tilde{h}_{\alpha,n,0}$ for sample size n and Hazard Rate α starting at Sample 3, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
9	10	3.663	4.370	4.928	5.512	6.339	7.008
9	11	3.242	3.909	4.424	4.956	5.695	6.289
9	12	3.037	3.678	4.167	4.666	5.347	5.886
9	13	2.910	3.530	3.999	4.471	5.108	5.608
9	14	2.821	3.425	3.875	4.325	4.927	5.393
9	15	2.755	3.344	3.780	4.213	4.787	5.228
9	16	2.704	3.280	3.704	4.121	4.673	5.089
9	17	2.662	3.228	3.641	4.046	4.578	4.980
9	18	2.628	3.184	3.587	3.982	4.495	4.886
9	19	2.599	3.147	3.542	3.927	4.425	4.801
9	20	2.575	3.115	3.503	3.879	4.365	4.727
9	21	2.554	3.086	3.468	3.837	4.314	4.664
9	22	2.535	3.061	3.438	3.800	4.265	4.610
9	23	2.519	3.039	3.410	3.767	4.224	4.563
9	24	2.504	3.018	3.386	3.737	4.186	4.516
9	25	2.490	3.001	3.363	3.710	4.152	4.477
9	26	2.479	2.985	3.344	3.686	4.121	4.441
9	27	2.468	2.970	3.326	3.664	4.093	4.408
9	28	2.458	2.957	3.309	3.643	4.067	4.376
9	29	2.449	2.944	3.293	3.624	4.043	4.349
9	30	2.441	2.933	3.279	3.608	4.021	4.322
9	35	2.408	2.887	3.223	3.539	3.933	4.219
9	40	2.385	2.855	3.183	3.490	3.873	4.149
9	45	2.367	2.831	3.152	3.453	3.825	4.093
9	50	2.354	2.812	3.129	3.425	3.791	4.051
9	60	2.334	2.784	3.094	3.383	3.737	3.988
9	70	2.322	2.765	3.070	3.353	3.701	3.948
9	80	2.310	2.751	3.053	3.333	3.675	3.917
9	90	2.305	2.741	3.040	3.316	3.655	3.894
9	100	2.301	2.732	3.028	3.304	3.638	3.874
9	125		2.716	3.010	3.281	3.610	3.841
9	150		2.708	2.998	3.266	3.592	3.822
9	175		2.701	2.989	3.254	3.578	3.805
9	200		2.700	2.983	3.248	3.569	3.794
9	250		2.691	2.974	3.237	3.555	3.780
9	300			2.967	3.229	3.546	3.769
9	350			2.967	3.225	3.541	3.761
9	400			2.966	3.223	3.536	3.757
9	450			2.973	3.220	3.532	3.751
9	500				3.217	3.530	3.748
9	600				3.215	3.526	3.744
9	700				3.217	3.522	3.739
9	800					3.520	3.735
9	900					3.521	3.735
9	1000					3.518	3.734

Table 2: Cutoffs $\tilde{h}_{\alpha,n,0}$ for sample size n and Hazard Rate α starting at Sample 10, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
14	15	3.378	3.909	4.308	4.710	5.253	5.679
14	16	3.049	3.579	3.972	4.366	4.890	5.293
14	17	2.886	3.415	3.805	4.191	4.702	5.091
14	18	2.784	3.311	3.697	4.076	4.574	4.954
14	19	2.713	3.238	3.619	3.993	4.479	4.847
14	20	2.661	3.182	3.560	3.927	4.403	4.759
14	21	2.620	3.138	3.511	3.873	4.341	4.688
14	22	2.588	3.103	3.471	3.828	4.287	4.627
14	23	2.561	3.072	3.437	3.789	4.241	4.577
14	24	2.538	3.045	3.408	3.755	4.199	4.527
14	25	2.519	3.023	3.381	3.725	4.163	4.485
14	26	2.503	3.003	3.359	3.698	4.131	4.449
14	27	2.488	2.986	3.338	3.674	4.101	4.415
14	28	2.476	2.970	3.320	3.652	4.074	4.381
14	29	2.464	2.956	3.303	3.632	4.048	4.353
14	30	2.454	2.944	3.288	3.615	4.026	4.325
14	35	2.415	2.893	3.227	3.542	3.936	4.221
14	40	2.389	2.859	3.185	3.492	3.874	4.150
14	45	2.370	2.833	3.154	3.455	3.826	4.094
14	50	2.356	2.813	3.130	3.426	3.792	4.052
14	60	2.335	2.786	3.095	3.383	3.737	3.988
14	70	2.322	2.766	3.071	3.354	3.701	3.949
14	80	2.311	2.752	3.053	3.333	3.675	3.917
14	90	2.305	2.741	3.040	3.316	3.655	3.894
14	100	2.301	2.732	3.028	3.304	3.638	3.874
14	125		2.716	3.010	3.281	3.611	3.841
14	150		2.708	2.998	3.266	3.592	3.822
14	175		2.702	2.989	3.254	3.578	3.805
14	200		2.700	2.983	3.248	3.569	3.793
14	250		2.692	2.974	3.237	3.556	3.779
14	300			2.968	3.230	3.546	3.769
14	350			2.967	3.226	3.541	3.761
14	400			2.965	3.222	3.536	3.757
14	450			2.968	3.220	3.532	3.751
14	500				3.217	3.530	3.748
14	600				3.214	3.526	3.744
14	700				3.217	3.522	3.739
14	800				3.218	3.520	3.736
14	900					3.521	3.736
14	1000					3.518	3.734

Table 3: Cutoffs $\tilde{h}_{\alpha,n,0}$ for sample size n and Hazard Rate α starting at Sample 15, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
20	21	3.261	3.713	4.046	4.373	4.803	5.129
20	22	2.969	3.432	3.769	4.099	4.529	4.851
20	23	2.819	3.291	3.630	3.961	4.390	4.712
20	24	2.725	3.201	3.542	3.873	4.298	4.616
20	25	2.660	3.138	3.480	3.810	4.234	4.548
20	26	2.612	3.091	3.433	3.762	4.184	4.495
20	27	2.574	3.055	3.397	3.724	4.141	4.449
20	28	2.545	3.025	3.366	3.691	4.105	4.409
20	29	2.520	3.001	3.340	3.664	4.074	4.375
20	30	2.501	2.981	3.318	3.641	4.047	4.343
20	35	2.436	2.909	3.241	3.554	3.945	4.229
20	40	2.400	2.868	3.193	3.499	3.879	4.154
20	45	2.377	2.839	3.159	3.459	3.829	4.097
20	50	2.360	2.817	3.133	3.428	3.794	4.054
20	60	2.337	2.788	3.097	3.385	3.738	3.989
20	70	2.323	2.767	3.072	3.354	3.702	3.949
20	80	2.312	2.753	3.054	3.334	3.675	3.917
20	90	2.306	2.742	3.040	3.316	3.655	3.894
20	100	2.301	2.733	3.029	3.304	3.638	3.874
20	125		2.717	3.010	3.281	3.611	3.841
20	150		2.709	2.998	3.266	3.593	3.822
20	175		2.702	2.990	3.255	3.578	3.805
20	200		2.699	2.983	3.248	3.569	3.794
20	250		2.691	2.974	3.237	3.556	3.779
20	300			2.968	3.230	3.546	3.769
20	350			2.967	3.225	3.541	3.761
20	400			2.965	3.223	3.536	3.757
20	450			2.968	3.220	3.532	3.751
20	500				3.218	3.530	3.747
20	600				3.215	3.526	3.744
20	700				3.217	3.522	3.739
20	800				3.220	3.519	3.735
20	900					3.521	3.736
20	1000					3.519	3.735

Table 4: Cutoffs $\tilde{h}_{\alpha,n,0}$ for sample size n and Hazard Rate α starting at Sample 21, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
7	8	2.448	3.143	3.709	4.322	5.210	5.949
7	9	2.371	3.075	3.630	4.219	5.053	5.747
7	10	2.333	3.025	3.561	4.119	4.890	5.516
7	11	2.310	2.979	3.499	4.025	4.741	5.317
7	12	2.285	2.940	3.435	3.937	4.617	5.149
7	13	2.262	2.902	3.382	3.862	4.509	5.006
7	14	2.241	2.867	3.334	3.794	4.409	4.888
7	15	2.222	2.839	3.293	3.739	4.326	4.777
7	16	2.205	2.812	3.257	3.691	4.257	4.681
7	17	2.189	2.788	3.224	3.648	4.197	4.608
7	18	2.175	2.766	3.196	3.610	4.141	4.537
7	19	2.162	2.747	3.170	3.574	4.092	4.477
7	20	2.151	2.730	3.147	3.544	4.050	4.427
7	21	2.140	2.715	3.125	3.519	4.013	4.381
7	22	2.130	2.699	3.104	3.491	3.978	4.336
7	23	2.120	2.687	3.087	3.467	3.947	4.298
7	24	2.111	2.673	3.069	3.447	3.919	4.264
7	25	2.105	2.663	3.055	3.426	3.894	4.233
7	26	2.096	2.652	3.041	3.409	3.871	4.206
7	27	2.089	2.641	3.028	3.392	3.849	4.177
7	28	2.084	2.632	3.017	3.378	3.829	4.154
7	29	2.078	2.624	3.006	3.364	3.810	4.129
7	30	2.073	2.615	2.996	3.352	3.795	4.110
7	35	2.049	2.583	2.952	3.297	3.721	4.025
7	40	2.033	2.559	2.920	3.257	3.673	3.964
7	45	2.019	2.539	2.898	3.228	3.636	3.919
7	50	2.011	2.524	2.879	3.205	3.605	3.884
7	60	1.997	2.503	2.849	3.168	3.557	3.829
7	70	1.984	2.487	2.830	3.142	3.524	3.786
7	80	1.979	2.476	2.814	3.126	3.500	3.761
7	90	1.971	2.466	2.804	3.112	3.485	3.742
7	100	1.969	2.461	2.794	3.101	3.469	3.723
7	125		2.450	2.782	3.084	3.446	3.696
7	150		2.441	2.770	3.071	3.431	3.677
7	175		2.435	2.762	3.061	3.418	3.658
7	200		2.429	2.755	3.053	3.409	3.650
7	250			2.747	3.044	3.394	3.636
7	300			2.748	3.035	3.385	3.628
7	350			2.740	3.034	3.383	3.621
7	400				3.026	3.378	3.618
7	450				3.026	3.376	3.614
7	500				3.021	3.373	3.613
7	600				3.026	3.366	3.603
7	700					3.363	3.600
7	800					3.359	3.603
7	900					3.363	3.595

Table 5: Cutoffs $\tilde{h}_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 8, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
9	10	2.861	3.526	4.045	4.588	5.348	5.948
9	11	2.561	3.211	3.710	4.220	4.929	5.485
9	12	2.433	3.069	3.553	4.042	4.710	5.225
9	13	2.362	2.986	3.455	3.926	4.564	5.051
9	14	2.312	2.925	3.383	3.837	4.443	4.915
9	15	2.275	2.880	3.327	3.768	4.348	4.795
9	16	2.245	2.844	3.283	3.712	4.274	4.694
9	17	2.222	2.813	3.244	3.665	4.208	4.617
9	18	2.201	2.786	3.212	3.623	4.150	4.544
9	19	2.184	2.763	3.183	3.584	4.099	4.482
9	20	2.168	2.743	3.157	3.552	4.056	4.431
9	21	2.155	2.726	3.133	3.525	4.018	4.384
9	22	2.142	2.709	3.111	3.497	3.982	4.338
9	23	2.131	2.695	3.093	3.473	3.950	4.301
9	24	2.121	2.680	3.074	3.451	3.922	4.267
9	25	2.113	2.669	3.060	3.430	3.897	4.235
9	26	2.104	2.657	3.044	3.412	3.873	4.208
9	27	2.096	2.646	3.032	3.395	3.851	4.179
9	28	2.090	2.637	3.020	3.380	3.831	4.155
9	29	2.084	2.628	3.009	3.367	3.812	4.130
9	30	2.079	2.619	2.998	3.354	3.797	4.111
9	35	2.052	2.585	2.954	3.298	3.722	4.025
9	40	2.035	2.561	2.921	3.258	3.673	3.964
9	45	2.020	2.540	2.899	3.229	3.637	3.920
9	50	2.012	2.525	2.879	3.205	3.606	3.884
9	60	1.998	2.504	2.850	3.169	3.557	3.829
9	70	1.985	2.488	2.831	3.143	3.524	3.787
9	80	1.980	2.477	2.814	3.126	3.501	3.761
9	90	1.969	2.467	2.804	3.112	3.485	3.742
9	100	1.969	2.462	2.794	3.101	3.469	3.722
9	125		2.451	2.782	3.084	3.446	3.696
9	150		2.441	2.770	3.071	3.431	3.677
9	175		2.435	2.762	3.061	3.418	3.658
9	200		2.428	2.755	3.053	3.409	3.650
9	250			2.747	3.045	3.394	3.636
9	300			2.748	3.035	3.385	3.628
9	350			2.742	3.035	3.383	3.621
9	400				3.026	3.378	3.618
9	450				3.026	3.376	3.614
9	500				3.021	3.372	3.613
9	600				3.027	3.366	3.603
9	700					3.362	3.600
9	800					3.359	3.603
9	900					3.363	3.595
9	1000					3.356	3.602

Table 6: Cutoffs $\tilde{h}_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 10, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
14	15	3.011	3.549	3.949	4.353	4.893	5.307
14	16	2.681	3.224	3.625	4.022	4.544	4.947
14	17	2.519	3.065	3.464	3.856	4.371	4.764
14	18	2.420	2.967	3.366	3.753	4.260	4.637
14	19	2.351	2.898	3.296	3.677	4.177	4.546
14	20	2.301	2.848	3.244	3.624	4.114	4.477
14	21	2.262	2.808	3.201	3.580	4.061	4.419
14	22	2.231	2.776	3.165	3.540	4.017	4.368
14	23	2.204	2.750	3.137	3.508	3.977	4.325
14	24	2.182	2.727	3.112	3.481	3.945	4.284
14	25	2.165	2.708	3.091	3.456	3.915	4.249
14	26	2.148	2.690	3.071	3.433	3.887	4.219
14	27	2.135	2.675	3.054	3.413	3.863	4.189
14	28	2.123	2.661	3.040	3.396	3.842	4.164
14	29	2.113	2.650	3.026	3.380	3.822	4.138
14	30	2.104	2.638	3.014	3.366	3.806	4.117
14	35	2.068	2.597	2.963	3.306	3.727	4.029
14	40	2.044	2.568	2.926	3.262	3.676	3.967
14	45	2.027	2.545	2.903	3.232	3.638	3.921
14	50	2.017	2.528	2.882	3.207	3.607	3.885
14	60	2.001	2.506	2.851	3.170	3.558	3.830
14	70	1.987	2.490	2.832	3.143	3.525	3.787
14	80	1.980	2.478	2.815	3.127	3.501	3.762
14	90	1.971	2.468	2.805	3.113	3.485	3.742
14	100	1.969	2.462	2.794	3.101	3.469	3.723
14	125		2.451	2.782	3.085	3.446	3.696
14	150		2.441	2.770	3.072	3.431	3.677
14	175		2.435	2.762	3.061	3.418	3.659
14	200		2.428	2.755	3.053	3.409	3.650
14	250			2.748	3.045	3.394	3.636
14	300			2.749	3.035	3.385	3.628
14	350			2.742	3.036	3.383	3.621
14	400				3.027	3.378	3.617
14	450				3.026	3.376	3.614
14	500				3.022	3.372	3.613
14	600				3.026	3.366	3.603
14	700					3.362	3.601
14	800					3.360	3.602
14	900					3.363	3.595
14	1000					3.356	3.602

Table 7: Cutoffs $\tilde{h}_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 15, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
20	21	3.031	3.496	3.835	4.166	4.603	4.937
20	22	2.718	3.198	3.545	3.884	4.318	4.647
20	23	2.556	3.046	3.398	3.739	4.174	4.505
20	24	2.451	2.949	3.303	3.647	4.082	4.409
20	25	2.378	2.880	3.238	3.582	4.019	4.339
20	26	2.323	2.829	3.187	3.531	3.966	4.287
20	27	2.279	2.787	3.148	3.492	3.925	4.242
20	28	2.243	2.754	3.116	3.459	3.890	4.206
20	29	2.214	2.728	3.090	3.433	3.861	4.172
20	30	2.190	2.704	3.068	3.410	3.838	4.147
20	35	2.111	2.629	2.989	3.327	3.744	4.041
20	40	2.070	2.586	2.942	3.275	3.685	3.974
20	45	2.043	2.557	2.912	3.240	3.644	3.925
20	50	2.027	2.536	2.889	3.213	3.611	3.888
20	60	2.007	2.510	2.855	3.173	3.559	3.831
20	70	1.992	2.492	2.834	3.145	3.526	3.788
20	80	1.982	2.479	2.816	3.128	3.502	3.762
20	90	1.972	2.469	2.806	3.114	3.485	3.742
20	100	1.969	2.463	2.795	3.102	3.470	3.723
20	125		2.452	2.782	3.085	3.446	3.696
20	150		2.441	2.770	3.071	3.431	3.677
20	175		2.436	2.762	3.061	3.418	3.658
20	200		2.425	2.756	3.054	3.409	3.650
20	250			2.746	3.044	3.394	3.636
20	300			2.749	3.035	3.385	3.629
20	350			2.742	3.036	3.383	3.621
20	400				3.027	3.378	3.618
20	450				3.025	3.377	3.614
20	500				3.022	3.373	3.613
20	600				3.029	3.366	3.603
20	700					3.363	3.601
20	800					3.360	3.603
20	900					3.363	3.595
20	1000					3.358	3.603

Table 8: Cutoffs $\tilde{h}_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 21, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
13	14	2.179	2.680	3.052	3.427	3.926	4.307
13	15	2.056	2.588	2.979	3.365	3.877	4.272
13	16	2.001	2.547	2.946	3.341	3.854	4.245
13	17	1.971	2.524	2.927	3.321	3.835	4.220
13	18	1.953	2.508	2.913	3.305	3.814	4.196
13	19	1.943	2.496	2.899	3.289	3.794	4.168
13	20	1.934	2.487	2.889	3.274	3.770	4.140
13	21	1.925	2.477	2.876	3.260	3.747	4.117
13	22	1.915	2.469	2.866	3.246	3.731	4.085
13	23	1.906	2.460	2.856	3.236	3.713	4.063
13	24	1.898	2.452	2.846	3.222	3.698	4.040
13	25	1.891	2.443	2.836	3.209	3.680	4.020
13	26	1.884	2.436	2.828	3.199	3.662	3.999
13	27	1.878	2.430	2.821	3.189	3.651	3.980
13	28	1.873	2.424	2.813	3.178	3.636	3.964
13	29	1.867	2.417	2.805	3.171	3.626	3.949
13	30	1.863	2.411	2.798	3.161	3.612	3.933
13	35	1.842	2.388	2.768	3.123	3.560	3.869
13	40	1.828	2.369	2.744	3.093	3.521	3.820
13	45	1.813	2.354	2.725	3.070	3.489	3.785
13	50	1.805	2.340	2.709	3.051	3.464	3.753
13	60	1.790	2.320	2.685	3.021	3.424	3.706
13	70	1.783	2.307	2.668	3.000	3.396	3.673
13	80	1.779	2.297	2.657	2.986	3.379	3.650
13	90	1.770	2.289	2.646	2.970	3.362	3.633
13	100	1.763	2.282	2.638	2.963	3.348	3.614
13	125		2.270	2.623	2.944	3.326	3.591
13	150		2.266	2.614	2.934	3.313	3.573
13	175		2.260	2.608	2.925	3.300	3.560
13	200		2.259	2.604	2.919	3.296	3.553
13	250			2.594	2.910	3.282	3.537
13	300			2.590	2.906	3.275	3.524
13	350			2.588	2.897	3.272	3.525
13	400				2.895	3.267	3.518
13	450				2.894	3.262	3.515
13	500				2.895	3.260	3.512
13	600				2.892	3.258	3.508
13	700					3.256	3.505
13	800					3.253	3.499
13	900					3.251	3.500
13	1000					3.245	3.497

Table 9: Cutoffs $\tilde{h}_{\alpha,n,6}$ for sample size n and Hazard Rate α starting at Sample 14, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
14	15	2.387	2.895	3.271	3.645	4.148	4.531
14	16	2.171	2.698	3.085	3.468	3.973	4.352
14	17	2.075	2.615	3.008	3.395	3.900	4.279
14	18	2.023	2.569	2.966	3.351	3.854	4.232
14	19	1.992	2.539	2.937	3.321	3.821	4.190
14	20	1.970	2.519	2.916	3.297	3.790	4.157
14	21	1.955	2.503	2.897	3.277	3.762	4.130
14	22	1.940	2.489	2.883	3.260	3.742	4.094
14	23	1.927	2.477	2.870	3.247	3.722	4.072
14	24	1.916	2.465	2.857	3.232	3.705	4.046
14	25	1.906	2.454	2.845	3.217	3.685	4.025
14	26	1.897	2.446	2.835	3.206	3.667	4.003
14	27	1.889	2.438	2.828	3.194	3.655	3.984
14	28	1.883	2.431	2.819	3.182	3.639	3.967
14	29	1.876	2.424	2.810	3.175	3.628	3.952
14	30	1.870	2.417	2.802	3.165	3.615	3.935
14	35	1.846	2.391	2.771	3.125	3.561	3.871
14	40	1.830	2.371	2.746	3.094	3.522	3.821
14	45	1.815	2.355	2.726	3.071	3.489	3.785
14	50	1.806	2.341	2.710	3.051	3.464	3.754
14	60	1.791	2.321	2.686	3.021	3.425	3.706
14	70	1.782	2.307	2.669	3.001	3.396	3.673
14	80	1.779	2.297	2.657	2.986	3.379	3.650
14	90	1.770	2.290	2.647	2.970	3.362	3.633
14	100	1.763	2.282	2.638	2.963	3.348	3.614
14	125		2.270	2.623	2.944	3.326	3.591
14	150		2.265	2.614	2.934	3.313	3.573
14	175		2.261	2.608	2.925	3.300	3.560
14	200		2.259	2.604	2.919	3.296	3.553
14	250			2.594	2.910	3.282	3.537
14	300			2.590	2.906	3.274	3.524
14	350			2.587	2.897	3.272	3.525
14	400				2.894	3.267	3.518
14	450				2.893	3.262	3.515
14	500				2.895	3.259	3.511
14	600				2.891	3.258	3.509
14	700					3.257	3.505
14	800					3.253	3.499
14	900					3.251	3.499
14	1000					3.245	3.497

Table 10: Cutoffs $\tilde{h}_{\alpha,n,6}$ for sample size n and Hazard Rate α starting at Sample 15, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
20	21	2.758	3.226	3.567	3.904	4.341	4.672
20	22	2.455	2.940	3.288	3.629	4.069	4.404
20	23	2.298	2.797	3.154	3.496	3.939	4.268
20	24	2.200	2.707	3.068	3.416	3.859	4.185
20	25	2.133	2.645	3.008	3.357	3.800	4.128
20	26	2.085	2.600	2.967	3.315	3.757	4.081
20	27	2.048	2.565	2.934	3.284	3.725	4.047
20	28	2.018	2.537	2.907	3.255	3.696	4.017
20	29	1.993	2.514	2.884	3.235	3.675	3.992
20	30	1.971	2.494	2.866	3.217	3.654	3.967
20	35	1.899	2.431	2.802	3.150	3.581	3.887
20	40	1.860	2.394	2.764	3.109	3.532	3.830
20	45	1.834	2.369	2.738	3.081	3.496	3.791
20	50	1.819	2.351	2.717	3.058	3.468	3.757
20	60	1.798	2.326	2.690	3.024	3.427	3.708
20	70	1.787	2.311	2.671	3.002	3.397	3.674
20	80	1.781	2.299	2.658	2.987	3.380	3.650
20	90	1.771	2.291	2.648	2.971	3.363	3.634
20	100	1.766	2.283	2.639	2.964	3.349	3.614
20	125		2.271	2.624	2.944	3.326	3.591
20	150		2.266	2.614	2.934	3.313	3.573
20	175		2.261	2.608	2.925	3.300	3.560
20	200		2.260	2.604	2.919	3.296	3.553
20	250			2.594	2.909	3.282	3.537
20	300			2.590	2.906	3.274	3.525
20	350			2.587	2.898	3.272	3.525
20	400				2.895	3.268	3.518
20	450				2.894	3.262	3.515
20	500				2.894	3.260	3.511
20	600				2.892	3.258	3.508
20	700					3.257	3.505
20	800					3.254	3.500
20	900					3.252	3.500
20	1000					3.245	3.497

Table 11: Cutoffs $\tilde{h}_{\alpha,n,6}$ for sample size n and Hazard Rate α starting at Sample 21, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
19	20	2.101	2.552	2.880	3.198	3.612	3.920
19	21	1.949	2.431	2.778	3.116	3.546	3.868
19	22	1.874	2.377	2.734	3.078	3.520	3.845
19	23	1.830	2.345	2.710	3.061	3.506	3.830
19	24	1.801	2.324	2.695	3.049	3.496	3.824
19	25	1.783	2.308	2.682	3.038	3.488	3.818
19	26	1.771	2.298	2.674	3.032	3.479	3.806
19	27	1.762	2.290	2.668	3.025	3.474	3.797
19	28	1.757	2.284	2.662	3.020	3.469	3.786
19	29	1.750	2.278	2.655	3.011	3.461	3.778
19	30	1.744	2.273	2.651	3.008	3.453	3.771
19	35	1.713	2.247	2.628	2.984	3.422	3.733
19	40	1.691	2.230	2.610	2.964	3.393	3.698
19	45	1.676	2.216	2.596	2.946	3.372	3.673
19	50	1.664	2.204	2.581	2.928	3.353	3.647
19	60	1.648	2.186	2.562	2.908	3.322	3.611
19	70	1.637	2.173	2.545	2.887	3.299	3.581
19	80	1.628	2.163	2.534	2.874	3.284	3.561
19	90	1.621	2.155	2.527	2.864	3.268	3.545
19	100	1.619	2.150	2.518	2.854	3.253	3.529
19	125		2.139	2.506	2.839	3.234	3.504
19	150		2.133	2.496	2.827	3.224	3.491
19	175		2.126	2.491	2.819	3.212	3.478
19	200		2.128	2.488	2.814	3.204	3.470
19	250			2.482	2.808	3.195	3.459
19	300			2.472	2.801	3.189	3.450
19	350			2.473	2.798	3.183	3.447
19	400				2.797	3.180	3.441
19	450				2.793	3.175	3.436
19	500				2.789	3.174	3.436
19	600				2.787	3.172	3.432
19	700					3.170	3.431
19	800					3.161	3.423
19	900					3.168	3.425
19	1000					3.166	3.419

Table 12: Cutoffs $\tilde{h}_{\alpha,n,9}$ for sample size n and Hazard Rate α starting at Sample 20, Parametric.

		α					
initial	n	0.05	0.02	0.01	0.005	0.002	0.001
20	21	2.273	2.728	3.060	3.379	3.798	4.106
20	22	2.050	2.532	2.877	3.209	3.636	3.954
20	23	1.945	2.444	2.800	3.141	3.576	3.895
20	24	1.883	2.394	2.757	3.103	3.541	3.868
20	25	1.843	2.361	2.728	3.078	3.521	3.848
20	26	1.816	2.338	2.709	3.062	3.503	3.826
20	27	1.797	2.322	2.695	3.048	3.493	3.812
20	28	1.785	2.309	2.684	3.039	3.484	3.798
20	29	1.775	2.299	2.673	3.027	3.473	3.788
20	30	1.765	2.291	2.666	3.021	3.463	3.779
20	35	1.725	2.257	2.636	2.991	3.426	3.737
20	40	1.698	2.235	2.614	2.967	3.396	3.700
20	45	1.680	2.219	2.598	2.948	3.374	3.675
20	50	1.667	2.206	2.583	2.930	3.354	3.648
20	60	1.649	2.188	2.562	2.909	3.322	3.612
20	70	1.638	2.174	2.546	2.888	3.299	3.582
20	80	1.628	2.163	2.535	2.874	3.284	3.562
20	90	1.622	2.155	2.527	2.865	3.268	3.545
20	100	1.619	2.150	2.518	2.854	3.253	3.529
20	125		2.139	2.506	2.839	3.234	3.504
20	150		2.133	2.496	2.827	3.224	3.491
20	175		2.126	2.492	2.819	3.212	3.478
20	200		2.128	2.488	2.814	3.204	3.470
20	250			2.482	2.808	3.195	3.459
20	300			2.472	2.801	3.189	3.450
20	350			2.473	2.798	3.183	3.447
20	400				2.797	3.180	3.441
20	450				2.793	3.175	3.436
20	500				2.789	3.174	3.436
20	600				2.788	3.172	3.432
20	700					3.170	3.431
20	800					3.161	3.424
20	900					3.168	3.425
20	1000					3.166	3.419

Table 13: Cutoffs $\tilde{h}_{\alpha,n,9}$ for sample size n and Hazard Rate α starting at Sample 21, Parametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
20	21	2.831	2.984	3.129	3.295	3.408	3.505
20	22	2.716	2.874	3.035	3.224	3.344	3.456
20	23	2.647	2.830	2.992	3.192	3.327	3.438
20	24	2.593	2.787	2.952	3.165	3.312	3.436
20	25	2.569	2.762	2.944	3.156	3.310	3.438
20	26	2.551	2.747	2.931	3.154	3.302	3.440
20	27	2.528	2.733	2.930	3.151	3.308	3.445
20	28	2.507	2.730	2.919	3.153	3.309	3.453
20	29	2.505	2.720	2.915	3.155	3.311	3.458
20	30	2.496	2.713	2.915	3.155	3.315	3.465
20	35	2.475	2.702	2.907	3.161	3.333	3.491
20	40	2.466	2.697	2.909	3.167	3.345	3.510
20	45	2.457	2.696	2.910	3.175	3.356	3.526
20	50	2.458	2.695	2.912	3.179	3.365	3.539
20	60	2.456	2.696	2.914	3.188	3.378	3.556
20	70	2.452	2.696	2.917	3.194	3.386	3.568
20	80	2.452	2.696	2.919	3.197	3.392	3.576
20	90	2.452	2.696	2.920	3.200	3.397	3.583
20	100	2.452	2.696	2.920	3.202	3.400	3.587
20	125		2.697	2.922	3.205	3.405	3.594
20	150		2.697	2.924	3.207	3.408	3.599
20	175		2.698	2.924	3.208	3.410	3.603
20	200		2.697	2.924	3.210	3.412	3.604
20	250		2.700	2.926	3.210	3.414	3.607
20	300		2.699	2.925	3.211	3.415	3.608
20	350		2.700	2.926	3.212	3.417	3.611
20	400		2.701	2.926	3.213	3.417	3.612
20	450		2.704	2.927	3.213	3.416	3.612
20	500			2.927	3.214	3.418	3.612
20	600			2.929	3.214	3.418	3.613
20	700			2.933	3.212	3.418	3.613

Table 14: Cutoffs $h_{\alpha,n,0}$ for sample size n and Hazard Rate α starting at Sample 21, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
14	15	2.698	2.847	2.946	3.065	3.182	3.239
14	16	2.595	2.781	2.911	3.092	3.163	3.275
14	17	2.531	2.730	2.917	3.074	3.184	3.317
14	18	2.515	2.717	2.901	3.083	3.204	3.322
14	19	2.500	2.702	2.890	3.067	3.228	3.342
14	20	2.474	2.704	2.888	3.087	3.233	3.365
14	21	2.470	2.687	2.888	3.098	3.248	3.364
14	22	2.465	2.703	2.884	3.097	3.250	3.381
14	23	2.454	2.680	2.893	3.099	3.270	3.399
14	24	2.452	2.690	2.880	3.109	3.270	3.403
14	25	2.446	2.673	2.890	3.117	3.273	3.418
14	26	2.438	2.687	2.894	3.124	3.284	3.427
14	27	2.438	2.682	2.885	3.134	3.294	3.441
14	28	2.429	2.686	2.893	3.138	3.303	3.450
14	29	2.428	2.681	2.898	3.139	3.304	3.456
14	30	2.432	2.684	2.903	3.145	3.310	3.465
14	35	2.427	2.683	2.903	3.159	3.334	3.494
14	40	2.424	2.682	2.908	3.168	3.347	3.513
14	45	2.419	2.676	2.911	3.175	3.359	3.529
14	50	2.417	2.682	2.914	3.182	3.369	3.542
14	60	2.416	2.684	2.916	3.189	3.380	3.559
14	70	2.414	2.682	2.919	3.195	3.388	3.571
14	80	2.412	2.684	2.919	3.199	3.393	3.576
14	90	2.412	2.683	2.920	3.200	3.399	3.586
14	100	2.414	2.683	2.921	3.203	3.402	3.585
14	125	2.411	2.684	2.923	3.206	3.406	3.597
14	150	2.414	2.685	2.924	3.206	3.408	3.599
14	175	2.412	2.686	2.924	3.209	3.411	3.603
14	200	2.412	2.683	2.923	3.211	3.412	3.604
14	250		2.683	2.925	3.211	3.413	3.607
14	300		2.682	2.925	3.209	3.417	3.608
14	350		2.686	2.925	3.213	3.415	3.611
14	400			2.926	3.213	3.417	3.612
14	450			2.927	3.213	3.416	3.610
14	500			2.927	3.216	3.419	3.613
14	600			2.935	3.216	3.418	3.613
14	700				3.215	3.417	3.613

Table 15: Cutoffs $h_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 15, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
20	21	2.831	2.984	3.129	3.298	3.409	3.507
20	22	2.705	2.878	3.041	3.233	3.349	3.467
20	23	2.615	2.833	3.003	3.203	3.338	3.451
20	24	2.580	2.797	2.959	3.180	3.324	3.447
20	25	2.547	2.774	2.958	3.168	3.322	3.446
20	26	2.519	2.752	2.941	3.160	3.317	3.449
20	27	2.508	2.737	2.933	3.159	3.316	3.453
20	28	2.494	2.732	2.929	3.157	3.317	3.463
20	29	2.481	2.719	2.924	3.163	3.321	3.467
20	30	2.476	2.716	2.921	3.163	3.324	3.473
20	35	2.448	2.698	2.915	3.167	3.339	3.499
20	40	2.434	2.690	2.918	3.174	3.351	3.516
20	45	2.431	2.688	2.914	3.178	3.362	3.531
20	50	2.424	2.686	2.918	3.184	3.371	3.544
20	60	2.419	2.686	2.918	3.190	3.381	3.560
20	70	2.416	2.683	2.921	3.196	3.389	3.571
20	80	2.413	2.685	2.920	3.199	3.394	3.576
20	90	2.413	2.684	2.921	3.200	3.399	3.586
20	100	2.414	2.684	2.921	3.203	3.402	3.585
20	125	2.412	2.684	2.923	3.206	3.406	3.597
20	150	2.415	2.685	2.924	3.206	3.408	3.599
20	175	2.411	2.686	2.925	3.209	3.412	3.603
20	200	2.412	2.684	2.923	3.210	3.412	3.604
20	250		2.683	2.925	3.211	3.413	3.607
20	300		2.682	2.925	3.209	3.417	3.608
20	350		2.686	2.925	3.212	3.415	3.610
20	400			2.926	3.213	3.416	3.612
20	450			2.926	3.213	3.416	3.610
20	500			2.927	3.216	3.419	3.613
20	600			2.935	3.217	3.418	3.612
20	700				3.216	3.418	3.613

Table 16: Cutoffs $h_{\alpha,n,3}$ for sample size n and Hazard Rate α starting at Sample 21, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
14	15	2.430	2.550	2.777	2.999	3.121	3.233
14	16	2.382	2.524	2.734	3.003	3.121	3.268
14	17	2.342	2.538	2.735	3.004	3.124	3.277
14	18	2.310	2.561	2.762	3.019	3.131	3.301
14	19	2.306	2.535	2.783	3.023	3.162	3.306
14	20	2.314	2.546	2.779	3.026	3.179	3.324
14	21	2.311	2.537	2.771	3.045	3.201	3.346
14	22	2.294	2.570	2.784	3.052	3.213	3.362
14	23	2.276	2.542	2.789	3.060	3.224	3.375
14	24	2.286	2.569	2.804	3.067	3.236	3.394
14	25	2.280	2.554	2.792	3.076	3.250	3.398
14	26	2.282	2.564	2.801	3.086	3.260	3.414
14	27	2.277	2.557	2.814	3.089	3.264	3.425
14	28	2.284	2.562	2.810	3.097	3.268	3.433
14	29	2.286	2.558	2.812	3.101	3.279	3.444
14	30	2.281	2.567	2.819	3.103	3.286	3.455
14	35	2.271	2.564	2.829	3.123	3.309	3.481
14	40	2.270	2.572	2.834	3.131	3.327	3.504
14	45	2.271	2.573	2.842	3.143	3.342	3.522
14	50	2.271	2.574	2.841	3.149	3.353	3.536
14	60	2.267	2.576	2.846	3.159	3.364	3.552
14	70	2.266	2.577	2.851	3.165	3.373	3.562
14	80	2.267	2.578	2.853	3.170	3.380	3.571
14	90	2.265	2.578	2.855	3.172	3.385	3.579
14	100	2.264	2.579	2.856	3.175	3.389	3.582
14	125	2.264	2.579	2.857	3.177	3.392	3.592
14	150	2.262	2.578	2.858	3.181	3.398	3.597
14	175	2.263	2.577	2.859	3.180	3.398	3.599
14	200	2.264	2.579	2.857	3.185	3.401	3.602
14	250		2.580	2.859	3.183	3.402	3.600
14	300		2.582	2.860	3.185	3.405	3.606
14	350		2.580	2.861	3.184	3.404	3.604
14	400			2.860	3.185	3.405	3.607
14	450			2.861	3.185	3.405	3.608
14	500			2.860	3.187	3.406	3.608
14	600			2.862	3.186	3.407	3.610
14	700				3.186	3.408	3.609

Table 17: Cutoffs $h_{\alpha,n,6}$ for sample size n and Hazard Rate α starting at Sample 15, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
20	21	2.734	2.910	3.066	3.268	3.393	3.501
20	22	2.573	2.785	2.974	3.192	3.328	3.450
20	23	2.494	2.720	2.919	3.154	3.302	3.436
20	24	2.448	2.694	2.891	3.138	3.294	3.435
20	25	2.415	2.663	2.883	3.130	3.284	3.439
20	26	2.394	2.637	2.870	3.123	3.288	3.441
20	27	2.373	2.626	2.864	3.122	3.294	3.448
20	28	2.354	2.626	2.848	3.118	3.299	3.452
20	29	2.343	2.613	2.850	3.117	3.302	3.458
20	30	2.330	2.611	2.855	3.120	3.301	3.463
20	35	2.306	2.594	2.844	3.133	3.320	3.487
20	40	2.291	2.583	2.841	3.139	3.332	3.507
20	45	2.284	2.583	2.847	3.148	3.345	3.524
20	50	2.277	2.583	2.845	3.152	3.355	3.537
20	60	2.270	2.580	2.848	3.160	3.365	3.552
20	70	2.268	2.578	2.852	3.167	3.374	3.562
20	80	2.268	2.579	2.854	3.170	3.380	3.571
20	90	2.266	2.578	2.855	3.173	3.386	3.580
20	100	2.265	2.579	2.856	3.176	3.390	3.582
20	125	2.264	2.579	2.857	3.177	3.392	3.591
20	150	2.263	2.578	2.858	3.181	3.398	3.597
20	175	2.263	2.577	2.859	3.180	3.398	3.599
20	200	2.263	2.579	2.857	3.185	3.401	3.602
20	250		2.580	2.858	3.182	3.402	3.600
20	300		2.583	2.861	3.185	3.405	3.606
20	350		2.580	2.861	3.184	3.404	3.603
20	400			2.860	3.185	3.405	3.606
20	450			2.862	3.185	3.406	3.608
20	500			2.860	3.187	3.405	3.608
20	600			2.861	3.186	3.408	3.609
20	700				3.186	3.408	3.609

Table 18: Cutoffs $h_{\alpha,n,6}$ for sample size n and Hazard Rate α starting at Sample 21, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
19	20	2.268	2.495	2.714	2.931	3.079	3.207
19	21	2.243	2.465	2.676	2.942	3.095	3.242
19	22	2.228	2.455	2.703	2.956	3.115	3.279
19	23	2.205	2.462	2.707	2.967	3.137	3.295
19	24	2.177	2.459	2.694	2.979	3.156	3.325
19	25	2.176	2.448	2.716	2.991	3.166	3.334
19	26	2.162	2.466	2.718	2.998	3.182	3.352
19	27	2.170	2.460	2.713	3.009	3.194	3.365
19	28	2.166	2.462	2.720	3.015	3.204	3.376
19	29	2.160	2.465	2.716	3.018	3.211	3.387
19	30	2.160	2.462	2.728	3.029	3.216	3.398
19	35	2.155	2.466	2.738	3.050	3.251	3.437
19	40	2.155	2.469	2.748	3.065	3.275	3.465
19	45	2.156	2.474	2.754	3.078	3.290	3.482
19	50	2.153	2.473	2.760	3.087	3.303	3.499
19	60	2.145	2.475	2.765	3.098	3.320	3.520
19	70	2.144	2.475	2.769	3.107	3.329	3.531
19	80	2.142	2.476	2.770	3.110	3.337	3.542
19	90	2.141	2.475	2.772	3.112	3.341	3.548
19	100	2.140	2.473	2.773	3.115	3.344	3.553
19	125	2.139	2.475	2.775	3.122	3.353	3.562
19	150	2.137	2.476	2.775	3.121	3.353	3.565
19	175	2.136	2.476	2.776	3.125	3.360	3.570
19	200	2.143	2.475	2.777	3.125	3.360	3.574
19	250		2.475	2.777	3.125	3.362	3.578
19	300		2.479	2.778	3.129	3.363	3.578
19	350		2.474	2.776	3.129	3.364	3.579
19	400			2.780	3.130	3.367	3.584
19	450			2.778	3.129	3.368	3.582
19	500			2.778	3.131	3.367	3.584
19	600			2.780	3.127	3.366	3.583
19	700				3.133	3.367	3.583

Table 19: Cutoffs $h_{\alpha,n,9}$ for sample size n and Hazard Rate α starting at Sample 15, Nonparametric.

		α					
initial	n	0.02	0.01	0.005	0.002	0.001	0.0005
20	21	2.395	2.605	2.817	3.030	3.178	3.318
20	22	2.308	2.533	2.769	2.993	3.166	3.306
20	23	2.265	2.494	2.729	2.981	3.163	3.314
20	24	2.229	2.479	2.742	2.987	3.166	3.334
20	25	2.219	2.475	2.720	2.997	3.178	3.342
20	26	2.211	2.473	2.727	3.004	3.193	3.357
20	27	2.197	2.470	2.718	3.011	3.205	3.371
20	28	2.188	2.470	2.732	3.019	3.209	3.382
20	29	2.184	2.476	2.734	3.022	3.218	3.392
20	30	2.175	2.470	2.732	3.034	3.221	3.401
20	35	2.159	2.471	2.739	3.053	3.254	3.438
20	40	2.155	2.470	2.749	3.066	3.276	3.465
20	45	2.156	2.476	2.755	3.079	3.290	3.483
20	50	2.154	2.474	2.760	3.087	3.303	3.499
20	60	2.146	2.476	2.765	3.098	3.320	3.520
20	70	2.145	2.475	2.769	3.107	3.330	3.531
20	80	2.142	2.476	2.770	3.110	3.337	3.542
20	90	2.142	2.475	2.772	3.112	3.341	3.548
20	100	2.140	2.474	2.773	3.115	3.344	3.553
20	125	2.139	2.475	2.775	3.122	3.353	3.562
20	150	2.136	2.476	2.775	3.121	3.353	3.565
20	175	2.137	2.476	2.776	3.125	3.360	3.570
20	200	2.143	2.475	2.777	3.125	3.360	3.574
20	250		2.475	2.777	3.125	3.362	3.578
20	300		2.479	2.778	3.129	3.363	3.578
20	350		2.474	2.776	3.129	3.364	3.579
20	400			2.780	3.130	3.367	3.584
20	450			2.778	3.130	3.368	3.582
20	500			2.778	3.131	3.368	3.584
20	600			2.780	3.127	3.366	3.583
20	700				3.134	3.367	3.583

Table 20: Cutoffs $h_{\alpha,n,9}$ for sample size n and Hazard Rate α starting at Sample 21, Nonparametric.