REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	MARCH 96	FINAL

4. TITLE AND SUBTITLE
TOPICS IN STATISTICAL PROCESS CONTROL

5. FUNDING NUMBERS
- NONE -

6. AUTHOR(S):
DAVID H. OLWELL, MAJOR, USA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
D/MATHEMATICAL SCIENCES,
USMA
WEST POINT, NY 10996

8. PERFORMING ORGANIZATION REPORT NUMBER
MADN-A-96-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
SAME AS #7

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
MADN-A-96-1

11. SUPPLEMENTARY NOTES

20

12a. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A: Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The inverse gaussian distribution can be used to model processes with skewed output. In this thesis, several Shewhard and CUSUM control schemes are developed for the inverse gaussian distribution. The behavior of these schemes is described. A new type of Shewhart chart, a self-starting Shewhart chart, is developed and applied to the inverse gaussian distribution. A second new type of control chart, which controls simultaneously for multiple parameters, is developed and shown to have some useful properties. Predictive Shewhart schemes, based on a diffuse prior, are developed and used for control.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Statistical process control; cusum; inverse gaussian distribution; JANUS; Loss exchange ratios	175
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

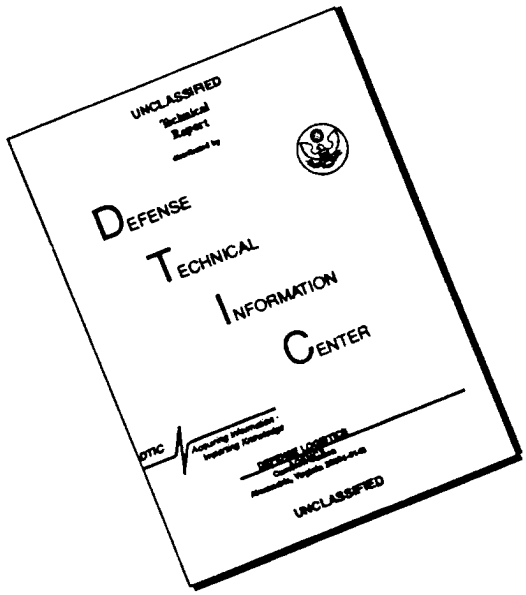
NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-01

Figure 5a

19960627 013

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Topics in Statistical Process Control

Technical Report MADN-A-96-1
Department of Mathematical Sciences
US Military Academy

David H. Olwell, Ph.D.[1]
Major, United States Army

March, 1996

[1]Department of Mathematical Sciences, US Military Academy, West Point, New York, 10996-1786

TOPICS IN STATISTICAL PROCESS CONTROL

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

DAVID HANLEY OLWELL

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

March 1996

Abstract

The inverse gaussian distribution can be used to model processes with skewed output. In this thesis, several Shewhart and CUSUM control schemes are developed for the inverse gaussian distribution. The behavior of these schemes is described. A new type of Shewhart chart, a self-starting Shewhart chart, is developed and applied to the inverse gaussian distribution. A second new type of control chart, which controls simultaneously for multiple parameters, is developed and shown to have some useful properties. Predictive Shewhart schemes, based on a diffuse prior, are developed and used for control.

We apply these methods to two applications. The first examines the control of the distribution of the output of a complex software package representing military combat, subject to continual revision. The second investigates the time to complete a task on a General Motors assembly line.

Acknowledgements

I would like to thank the faculty of the School of Statistics of the University of Minnesota. They have developed a true community of scholars, a community which encourages and supports its graduate students. I have learned a great deal about the art of teaching from them. Their enthusiasm and style were contagious.

I especially thank Professor Doug Hawkins, my thesis advisor, who has guided my studies with kindness, wisdom, wit, and patience since 1992. His encylopedic knowledge of applied statistics and his own extensive work in statistical process control were invaluable, and shared graciously. He suggested many revisions and additions to the work, which clarified and extended it. (Remaining errors, if any, are my responsibility.) He was sensitive to my military requirements, and made my work a priority when I needed his quick assistance. It is not hyperbole to say that I could not have accomplished this work with another advisor.

I thank my thesis reviewers, Professors Gary Oehlert and John Anderson, for their careful reading of the work, which resulted in many improvements.

I thank Professor Luke Tierney for freely sharing his excellent *Xlisp-Stat* software, used extensively in this thesis.

I thank my friends among the graduate students of the School for their help over the years as we studied together – particularly Marilyn Agin, Efstathia Bura, Francesca Chiaromonte, Wen-Lin Lai, Bret Musser, and Dave Nelson.

The staff of the School of Statistics has been genuinely helpful, and I thank them for their many kindnesses.

I have enjoyed my contacts with the other units of the University during my graduate schooling, particularly the School of Mathematics, where I earned a Master's

degree in 1989; the Department of Mechanical Engineering, where I earned a minor in 1989; the Carlson School of Business, where I studied quality management methods and theory in 1993 and 1994; and the Army High Performance Computing Research Center, which provided me a "home away from home", computing support, and administrative assistance.

I am grateful to Mr. Tom Herbert of the Rand Corporation for providing *gratis* one of the data sets analyzed in Chapter 7, and to Ms. Angela Stich, also of Rand, who compiled the files and assisted in the post-processing. Mr. David Durda, of the US Army Training and Doctrine Analysis Center, White Sands Missile Range, provided a second data set with post-processing.

Colonel Chris Arney and the faculty of the Department of Mathematical Sciences at West Point have been extremely supportive, allowing me a reduced teaching load as I finished this thesis.

My wife, Karen, and my children have been supportive these four years as I studied and was frequently absent. Karen has tolerated with humor and patience my preoccupied state while wrestling with the problems of this thesis. Karen also proofread four drafts of this thesis, with painstaking attention to detail. I thank her for all her assistance.

Finally, I will forever be grateful to Brigadier General (retired) Frank R. Giordano, my former teacher as a cadet and my boss during my first three year tour as an instructor at West Point. He gave me freedom to experiment. General Giordano accepted my mistakes and turned them into learning opportunities. He selected me for this advanced schooling, and has "set the example" of soldier–scholar for me. His love of West Point, love of mathematics, and love of cadets have been deeply inspirational to me, and his faith in me has changed my life.

Contents

List of Tables

List of Figures

Chapter 1

Problem statement

1.1 Control Charts

A control chart is a graphical tool for detecting departures from an assumed distribution. Charts are used to display information taken from samples from the ongoing process, and signal by various algorithms when it is likely that a model departure exists.

There are two broad types of control charts. The first type, due to Shewhart [Shewart, 1931], measures when individual samples of fixed size indicate a departure from the model. For example, a power spike occurs during the manufacturing process, causing the characteristic(s) of interest of one batch of output to follow a different distribution from historical patterns. Shewhart charts are simple, and signal large changes quickly.

The second type of control chart, due to Page [Page, 1954], detects small, persistent changes in the process output. Page introduced the Cumulative Sum Chart (CUSUM). Another variation on the idea is the Exponentially Weighted Moving Average (EWMA), due to Roberts [Roberts, 1959]. These charts use the previous history of the process and provide a quicker method of detecting small changes in the process.

One can chart different attributes of the process. Historically, one has charted measures of process location and dispersion.

Control charts have been used extensively in manufacturing industries, chemical industries, and business. The most common assume that the underlying process can be well modeled by the normal distribution.

Control chart methodologies have been developed for processes modeled by many members of the exponential family, including the normal, gamma, poisson, and binomial. In 1989, Edgeman proposed Shewhart control charts for the inverse gaussian (IG) distribution [Edgeman, 1989].

The Shewhart charts suffer from limitations. A signal for a change in location can mean either a change in process mean, or an increase in process variability. This is discussed below. Accordingly, one must consider both charts together when diagnosing shifts in process centrality. Furthermore, they are slow to signal small changes in the distribution.

We shall extend existing theory to include CUSUM schemes for inverse gaussian distributed random variables. We explore their properties. Later, we will propose new methods for constructing Shewhart charts, and explore their properties.

1.1.1 Shewhart charts

Many excellent references describe the basic Shewhart control scheme. They include Montgomery [1991] and Barnard [1959], among others. We survey the general approach.

From historical data or first principles, one models the distribution of the process for a given characteristic when it is "in-control". "In-control" means that the process is following the postulated distribution. Frequently, the process is taken to be normally distributed, but there are schemes for discrete distributions as well as for non-normal interval data. The process characteristic could be some physical dimension of the output of the process, or the number of defects, or whatever is of interest to those monitoring the process.

We draw samples from the process upon which to base our inference about the process characteristic. If we draw samples of size greater than one, we desire to

obtain a "rational subgroup". A rational subgroup is a sample which is expected to be homogenous. This is best illustrated by counter-example: a sample which spanned two shifts of workers would not be expected to be completely homogenous, because of worker to worker variation. A rational subgroup has no plausible *a priori* explanation for being a mixed distribution of in-control and out-of-control.

From the distribution of the characteristic, we determine the sampling distribution for our sample size. We then construct limits on acceptable values for sufficient statistics for draws from the sampling distribution, using either probability limits or a rule of thumb based on a given number of standard deviations. We construct the control chart by plotting the sufficient statistic vs. draw number on a chart which already has graphed the expected value of the statistic, the upper control limit, and the lower control limit. See Figure 1.1 for an example.

If the sample statistic is within the control limits, we take no action. If the statistic is outside the limits, we consider that an "out-of-control" signal. We reject the hypothesis that the process is still following the postulated model, and investigate.

Analogous with hypothesis testing, we are concerned with two types of errors. First, the process may signal a model departure when none exists. If the control limits are constructed so that the probability that the sample is within the limits is $1 - \alpha$, then the probability of a false signal is α. The number of samples to a false signal is called the Run Length. The distribution of the run length for a process in-control is geometric. The expected number of samples until a false signal is called the Average Run Length (ARL) and is equal to $\frac{1}{\alpha}$. This is the traditional measure of effectiveness for a control chart. Long ARL is desirable when in-control, for there are costs associated with investigating false signals.

Given a model departure of specified form and magnitude, we can compute the probability of a signal, and the corresponding ARL for detecting the model departure. When out-of-control, short ARLs are desirable, since there are (usually) costs

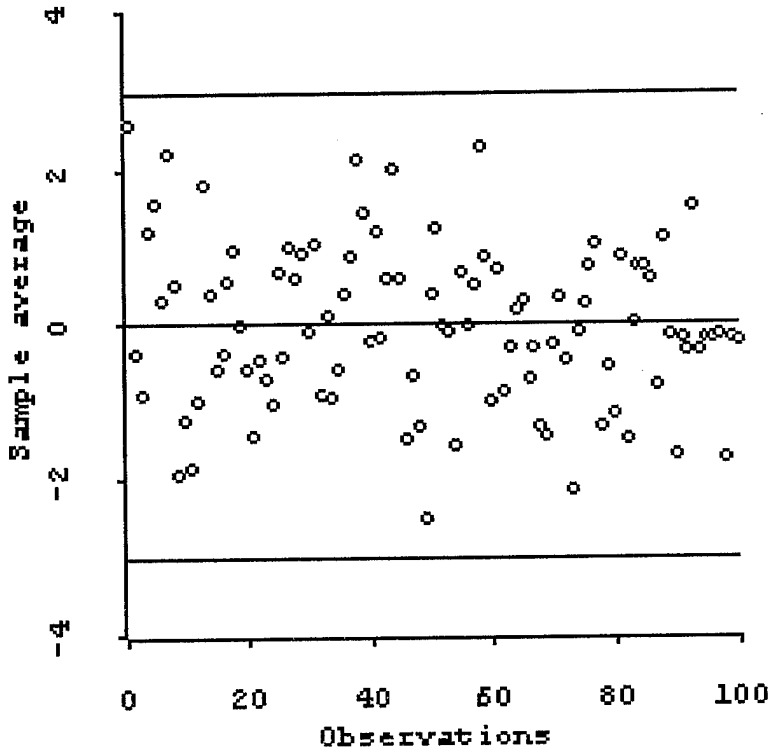


Figure 1.1: A Shewhart control chart for the sample average, with the process in-control.

associated with model departures of various sizes. We can construct the operating characteristic curve as in classical hypothesis testing.

It is accepted practice to chart for both location and scale, when dealing with normal processes. Consider a shift from the distribution $X_i \sim N(\mu, \sigma)$ to the distribution $X_o \sim N(\mu, \Delta \times \sigma^2)$, $\Delta > 1$. A chart for location alone would eventually detect such a shift, but it would not be efficient. Worse, such a signal might be interpreted as a mean shift, not a scale shift. This is illustrated in Figures 1.2 and 1.3. Notice both location and scale shifts are signaled on the location chart. The scale shift signals because the increased variability increases the probability that the process will exceed the control limits.

Conventional practice interprets signals as follows: if there is a signal on the scale chart, consider that a shift in scale has occurred. If there is a signal on the mean chart only, consider that a shift in mean has occurred. If there is a signal in both charts, it may be due to either a shift in scale alone, or a shift in both scale and location. Investigate both possibilities.

As with classic hypothesis testing, there is a tradeoff between the probability of type I and type II error, here given by the relative size of the ARL in-control vs. ARL out-of-control for various model departures.

Depending on our criteria, we select sample size, sample frequency, and the control limits to best meet our needs. The criteria can be economic, and incorporate different losses for false positive signals, false negative signals, and sampling costs. The criteria can be strictly statistical. Optimal design against these criteria is discussed by many authors, including Girshick and Rubin, [1952], Bather, [1963], Ross [1971], Savage[1962], and Taylor [1965, 1967]. A good survey is given in Montgomery [1991].

The advantages of Shewhart charts are simplicity and immediate sensitivity to large model departures. However, for small model departures, the ARL until a signal

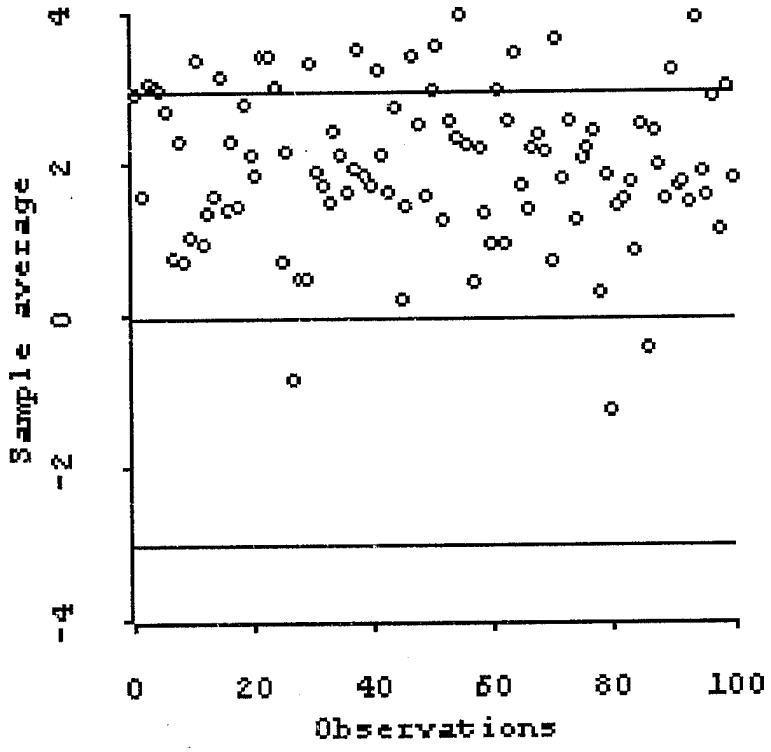


Figure 1.2: A Shewhart chart for the mean for a process out-of-control due to a location shift.

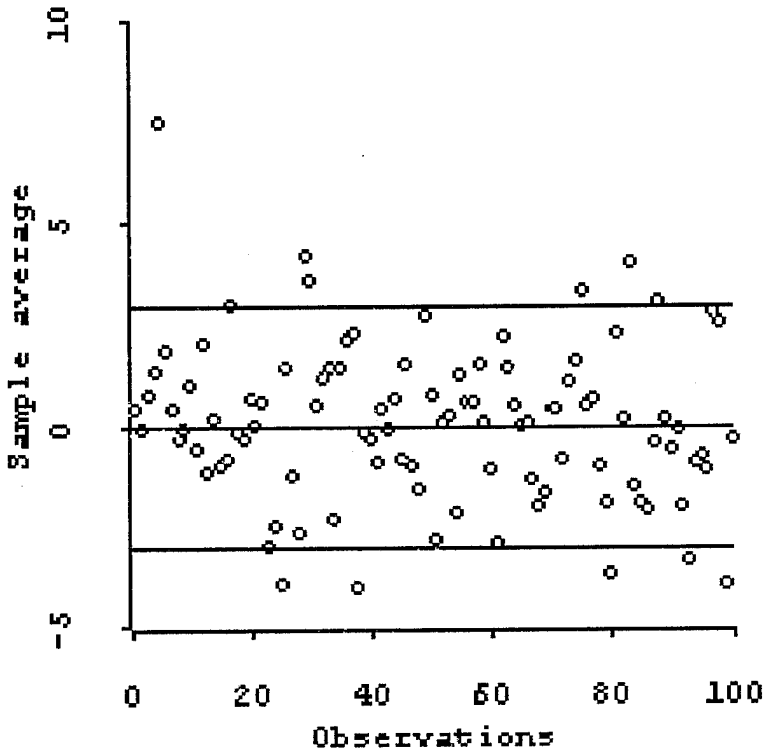


Figure 1.3: A Shewhart chart for the mean for a process out-of-control due to increased variability.

can be quite large. Depending on the losses associated with large out-of-control ARLs, this can be costly for the process manager.

1.1.2 Cumulative Sum (CUSUM) control charts

The Shewhart chart tests the hypothesis that the sample came from the "in-control" distribution. Treating each sample separately, it does not use all the information available in the case where the out-of-control affects more than one rational subgroup. Various rules for declaring signals that the process is out-of-control have been proposed to use more of the data available. For example, one author proposes the process be declared out-of-control if 8 points in sequence are above the centerline of the chart, reasoning that the probability that 8 consecutive independent draws from a distribution are above the median is so unlikely ($1/2^8 = .0039$) as to signal a departure from the null hypothesis [Western Electric, 1956]. More generally, the Western Electric rules signal the process out-of-control if k out of N of the preceding points are above or below the median. Champ and Woodall [1987, 1990] discuss supplementary rules and provide methods for determining ARLs.

The cumulative sum chart generalizes this idea. It works with the sum of previous observations (or transformations of observations.) As A.L. Goel says,

> This system of charting takes full advantage of the historical record and provides a rapid means of detecting shifts in the process level. [Goel, 1982]

In other words, the Shewhart chart is designed to detect large, isolated shifts in the process. The CUSUM is designed to detect persistent, perhaps small, shifts in the process.

Standards references for the CUSUM include Van Dobben de Bruyn [1966], Johnson [1961], and Johnson and Leone [1962a, 1962b, 1962c].

The roots of the CUSUM lay with the sequential probability ratio test (SPRT). This test was developed by Abraham Wald while at the Statistical Research Group at Columbia during the Second World War. It is interesting (to this candidate) that he attributes the genesis of the theory "in connection with some comments made by Captain G. L. Schuyler of the Bureau of Ordnance, Navy Department." [Wald, 1947].

Wald published on the theory of cumulative sums of random variables [Wald, 1944].

Page [Page, 1954] adapted these methods to quality control, proposing the first quality charts based on cumulative sums of observations and calling them "CUSUM" charts.

We will use the following decision interval characterization of the CUSUM. Consider the model where X has density $f(x|\theta)$, i.e. X is a random variable whose distribution is indexed by the parameter θ. We consider θ known and fixed at a value, say θ_0. We observe the sequence $X_1, X_2, \ldots, X_n$. We are interested in testing the hypothesis that $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$. Instead of dealing with a composite alternative hypothesis, we consider two point alternatives: $\theta = \theta_l$ or $\theta = \theta_u$.

Wald conjectured [Wald, 1947] and later proved [Wald and Wolfowitz, 1948] that the sequential probability ratio test was optimal for deciding between the two point hypotheses in the sense that the expected number of points sampled before a decision could be reached was minimized with the SPRT. A precise statement of these optimality properties of the SPRT in a decision framework can be found in [Ferguson, 1967].

The SPRT considers

$$\Lambda_n = \frac{f(X_1, X_2, \ldots, X_n|\theta_1)}{f(X_1, X_2, \ldots, X_n|\theta_0)} = \prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \tag{1.1}$$

where $f(x|\theta)$ is the joint or marginal density as appropriate. The SPRT accepts

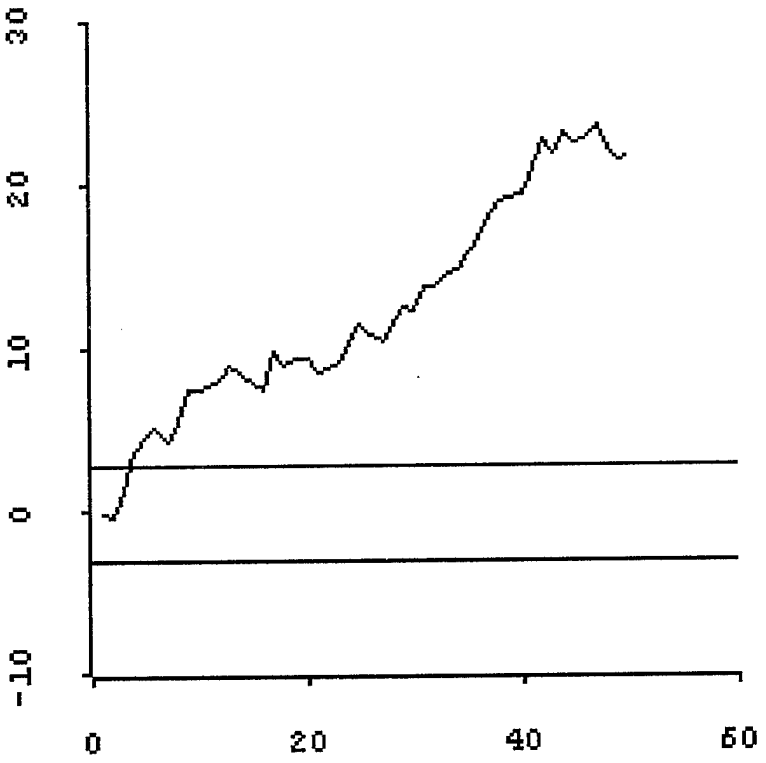


Figure 1.4: A graphical description of the SPRT.

$H_0 : \theta = \theta_0$ if $\Lambda_n \leq A$, accepts $H_a : \theta = \theta_1$ if $\Lambda_n \geq B$ and otherwise continues sampling. This is illustrated in Figure 1.4, with $A = -3$ and $B = 3$, where the null hypothesis would have been rejected at observation number 4.

In practice, we work with the log-likelihood, or $\ln(\Lambda_n)$, which results in a cumulative sum. We accept, reject, or continue sampling based on the value of this cumulative sum. As we have written it, the log-likelihood ratio will have a negative expected value when the process is in-control. When the process is well modeled by the alternate hypothesis, the log-likelihood ratio will have a positive expected

value. As a result, when the process is in-control, the sum tends downward. When the process is out-of-control at the alternative distribution, the sum tends upward. When the sum is above a certain limit, we have evidence in favor of the alternative hypothesis. When the sum is below a certain limit, we decide in favor of the null hypothesis. When the sum is in-between the limits, we continue to sample.

The CUSUM is based on this test, except that the **null hypothesis is never accepted.** Instead, we restart the test each time the evidence favors the null hypothesis. We consider the evidence as favoring the null hypothesis whenever the sum is negative. At that point, we start over by resetting the sum to zero. We sample until we reject the null hypothesis in favor of the point alternative.

For members of the one-parameter exponential family, we have the following general case:

$$\ln f(x|\theta) = a(x)b(\theta) + c(x) + d(\theta)$$

Then the log-likelihood ratio SPRT implies:

$$a \leq \ln \Lambda_n \leq b \tag{1.2}$$

$$a \leq \sum_{i=1}^{n} a(x_i)b(\theta_1) + c(x_i) + d(\theta_1) - (\sum_{i=1}^{n} a(x_i)b(\theta_0) + c(x_i) + d(\theta_0)) \leq b$$

$$\frac{a}{b(\theta_1) - b(\theta_0)} \leq \sum_{i=1}^{n} \left(a(x_i) + \frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} \right) \leq \frac{b}{b(\theta_1) - b(\theta_0)} \tag{1.3}$$

Accordingly, the SPRT is equivalent to testing if $a' < \sum a(x_i) + k < b'$, where

$$a' = \frac{a}{b(\theta_1) - b(\theta_0)} \tag{1.4}$$

$$b' = \frac{b}{b(\theta_1) - b(\theta_0)} \tag{1.5}$$

$$k = \frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)} \tag{1.6}$$

The test statistic in the SPRT is the cumulative sum, $\sum_{i=1}^{n}(a(x_i)+k)$.

The CUSUM differs from the SPRT is two ways. First, we never accept the null hypothesis, so anytime the CUSUM is negative, which favors the null hypothesis, we start over. In other words, the indices for the sum are different. Less importantly, the way we have constructed the SPRT rejects the null hypothesis if the CUSUM is greater than a' and $b(\theta_1) > b(\theta_0)$. We prefer this rejection region on the positive side. Reversing the roles of θ_0 and θ_1 if necessary accomplishes that for the case $b(\theta_1) < b(\theta_0)$. Since k is usually negative, we also change the sign of k in our notation. This notation is consistent with the industry standard, described in Van Dobben de Bruyn [1966].

The scheme then becomes:

$$S_0^+ = 0 \tag{1.7}$$

$$S_n^+ = \max(0, S_{n-1}^+ + a(x_n) - k^+) \tag{1.8}$$

and signals when $S_n^+ \geq h^+$.

Say in the case above, $\theta_1 > \theta_0$. We can also construct a scheme to detect the changes in the other direction, for $\theta_2 < \theta_0$. The second scheme, for a shift in the opposite direction, is:

$$S_0^- = 0 \tag{1.9}$$

$$S_n^- = \min(0, S_{n-1}^- + a(x_n) - k^-) \tag{1.10}$$

This chart signals if $S_N^- \leq h^-$. Many practitioners change the sign of k^- to be positive if it is negative, so one frequently sees Equation 1.10 written as $S_n^- = \min(0, S_{n-1}^- + a(x_n) + k^-)$.

We run both schemes simultaneously to detect changes in θ. This results in a pair of schemes being plotted on the same chart, as illustrated in Figure 1.5.

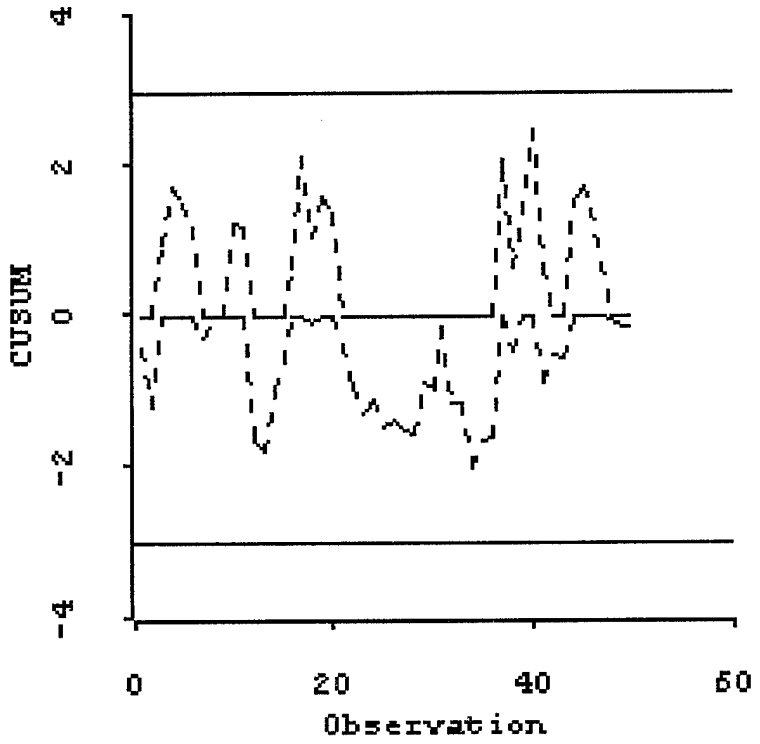


Figure 1.5: A CUSUM chart, using the decision interval format.

We can also decide, based on costs of being out-of-control in one direction or the other, to set asymmetric values of θ_1 and θ_2, resulting in asymmetric values of h^+ and h^-.

The upward and downward CUSUMs will each have its own ARL. If $k^+ + k^- > |h^+ - h^-|$, then overall ARL for the two-sided scheme is :

$$ARL = \frac{1}{\frac{1}{ARL_{lower}} + \frac{1}{ARL_{upper}}} \tag{1.11}$$

[Van Dobben de Bruyn, 1966].

With uniform scaling, it is possible to run both the Shewhart and the CUSUM schemes on the same chart. Some practitioners recommend putting as many as 6 charts on the same graph, for the two parameter normal distribution: A Shewhart chart for location, another Shewhart chart for scale, two CUSUMs for location (upper and lower), and two CUSUMs for scale (upper and lower). This is illustrated in Figure 1.6. About such multiple charts, Hawkins [1992b] says,

> Suprisingly, this chart with up to 6 things plotted does not contain much 'clutter'. Under control, each of the four cusums spend much of its time running along the axis, and the two positive and negative cusums are confined to their own halves of the paper. Showing them as solid lines and adding the Shewhart points as individual unconnected symbols generally gives quite a clear presentation.

Moustakides [1986] showed that the CUSUM scheme enjoyed the same optimality properties as the SPRT. Among all tests with the same in-control ARL, the CUSUM had the smallest expected run length out-of-control. The reader is referred to Moustakides for a precise statement and proof.

An interesting interpretation of the Shewhart chart classifies it as a CUSUM with $h = 0$ and $k =$ the control limit. Then the Shewhart chart is a CUSUM which

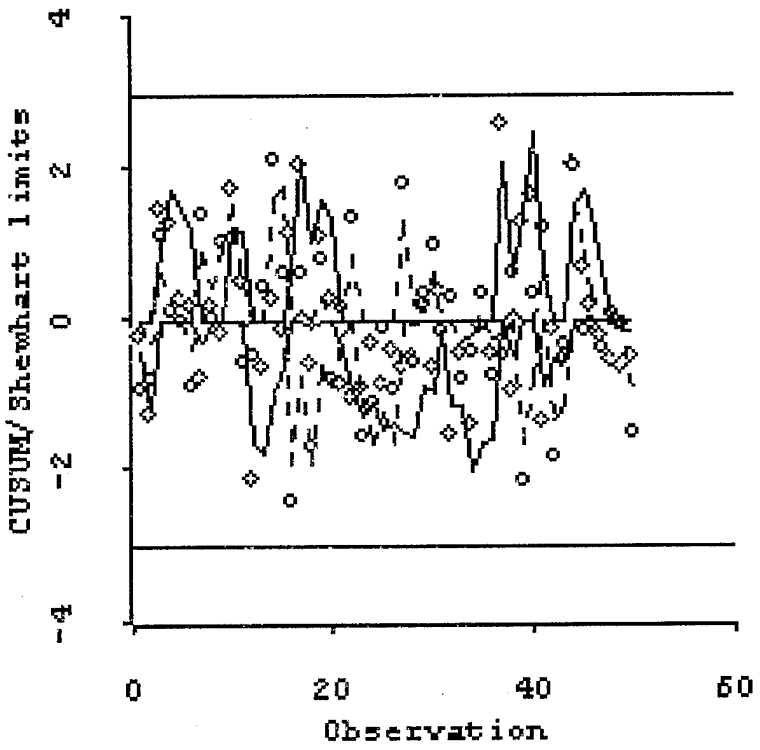


Figure 1.6: A multiple chart, showing both Shewhart charts for location and scale, and the 4 CUSUMs for location and scale.

signals immediately if $a(X) \geq k$, but restarts for any observation less than k.

We will examine both CUSUMs and Shewhart charts in the body of this thesis.

1.2 Inverse Gaussian processes

The inverse gaussian distribution has its genesis in the analysis of Brownian motion. Following Chhikara and Folks [1989], we characterize the Wiener process $X(t)$ with drift ν and variance σ^2 as follows:

1. $X(t)$ has independent increments; for $t_1 < t_2 < t_3 < t_4$, we have $X(t_2) - X(t_1)$ independent of $X(t_4) - X(t_3)$

2. $X(t_2) - X(t_1)$ is normally distributed with mean $\nu(t_2 - t_1)$ and variance $\sigma^2(t_2 - t_1)$, with $t_2 > t_1$.

Schroedinger [1915] first showed that the distribution of the first time until the process $X(t) > a$ for $\nu > 0$, $a > 0$ was inverse gaussian. See Figure 1.7 for an illustration.

This characterization of the inverse gaussian is useful for the applications which follow in this thesis. Many processes can be well modeled by Brownian motion with drift. We shall see how the attrition of forces in a military model can be approximately modeled by the inverse gaussian distribution. We shall also examine how well the time to complete the work at a station on an automobile assembly line is modeled by an inverse gaussian distribution.

The inverse gaussian distribution is also useful for modeling of positive, skewed processes in general, even when the underlying mechanics do not immediately suggest a theoretical basis for Brownian motion passage times.

The reader may notice a striking similarity between Figures 1.4 and 1.7. Wald [1944, 1945, 1947] showed that the distribution of the stopping times in the SPRT

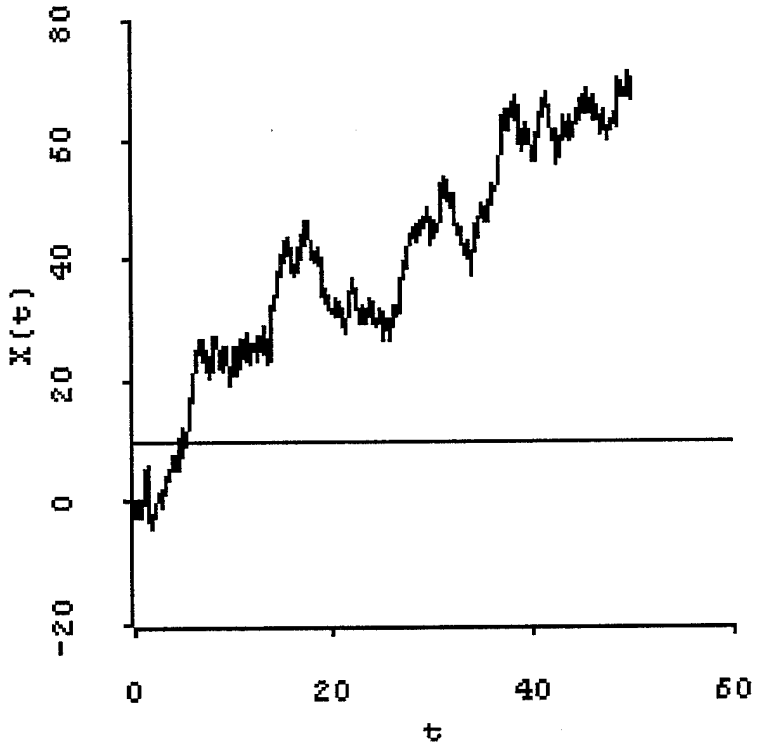


Figure 1.7: First passage time illustration for Brownian motion with drift

with lower limit $b = -\infty$ and upper limit a was approximately distributed as inverse gaussian. Since the stopping times for the SPRT are discrete random variables and the inverse gaussian is a continuous random variable, the two distributions can only be approximately equal.

1.2.1 Well Known Properties

In this section, we list several well known results about the inverse gaussian distribution, which later prove useful.

1.2.1.1 PDF

The probability density function (pdf) for $X \sim IG(\mu, \lambda)$ is

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi}}\, x^{-3/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0 \tag{1.12}$$

Several density curves for various values of μ and λ can be seen in Figures 1.8 and 1.9.

1.2.1.2 CDF

The cumulative distribution function for $X \sim IG(\mu, \lambda)$ is

$$F(x;\mu,\lambda) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right] + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right] \tag{1.13}$$

where $\Phi(x)$ is the CDF of the standard normal distribution. [Chhikara and Folks, 1989]

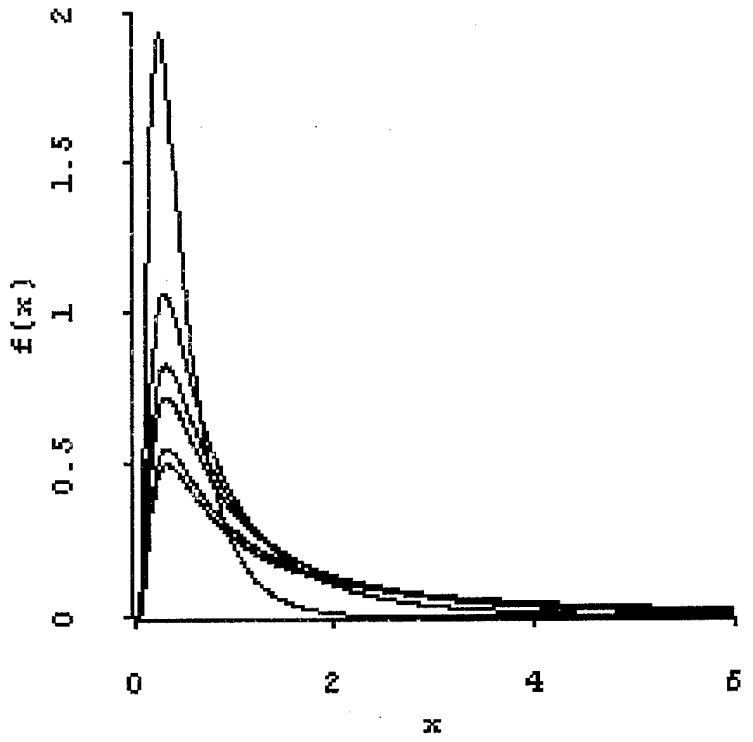


Figure 1.8: A sheaf of $IG(\mu, 1)$ densities for $\mu = .5, 1, 1.5, 2, 5,$ and 10

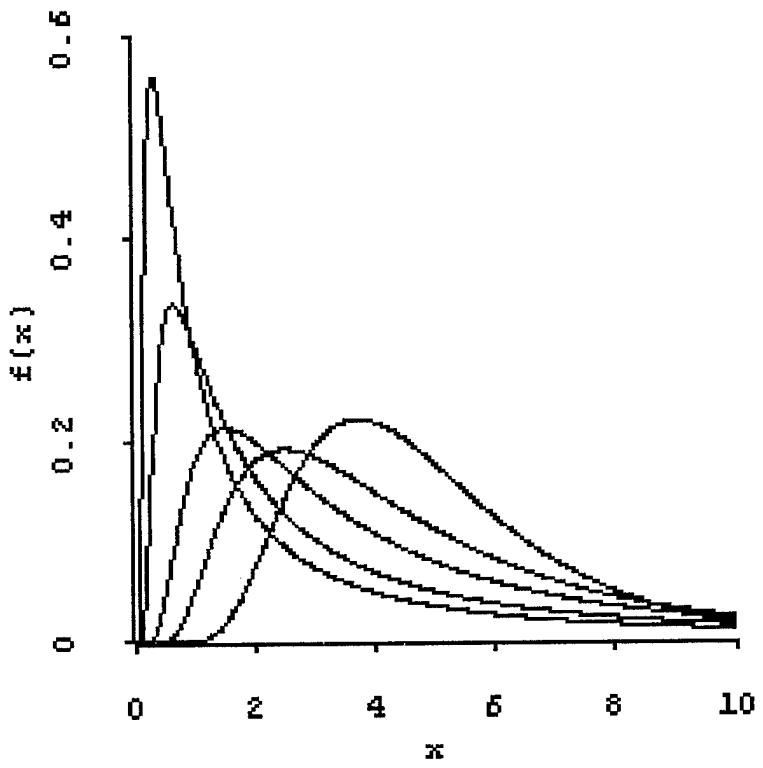


Figure 1.9: A sheaf of $IG(5, \lambda)$ densities for $\lambda = 1, 2, 5, 10$, and 25

1.2.1.3 First passage time interpretation

It is well known [Chhikara and Folks, 1989] that for a Wiener process with positive drift ($\nu > 0$) the first passage time to the barrier is inverse gaussian $IG(\mu, \lambda)$, with $\mu = \frac{(a-x_o)}{\nu}$ and $\lambda = \frac{(a-x_o)^2}{\sigma^2}$.

1.2.1.4 Characteristic function and moments

The characteristic function for $X \sim IG(\mu, \lambda)$, $\Psi(t) = E\exp(iXt)$, is

$$\Psi(t) = \exp\left(\frac{\lambda}{\mu}\left[1 - \sqrt{\left(1 - \frac{2i\mu^2 t}{\lambda}\right)}\right]\right).$$

From this, it follows that the mean of X is μ and the variance of X is $\frac{\mu^3}{\lambda}$.

1.2.1.5 Member of the exponential family

The inverse gaussian distribution is known to belong to the exponential family of order two. Let $\theta = \frac{\lambda}{\mu^2}$. Then the pdf can be expressed as

$$f(x; \lambda, \theta) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp(\lambda\theta/2)\, x^{-3/2} \exp\left(-\frac{1}{2}(\frac{\lambda}{x} + \theta\, x)\right) \tag{1.14}$$

which is of the natural exponential family

$$c(x)d(\Theta)\exp(a(x)\cdot b(\Theta))$$

with $b(\Theta) = \Theta = (\lambda, \theta)$ and $a(x) = -1/2(x^{-1}, x)$.

Accordingly, for a random sample $\mathbf{X}$ from the inverse gaussian, the two dimensional statistic $(\sum X, \sum X^{-1})$ is complete and minimal sufficient.

1.2.1.6 MLEs of parameters and their distribution

For a random sample $X_1, X_2, \ldots, X_n$ where $X_i \sim IG(\mu, \lambda)$, the likelihood function is

$$L(\mu,\, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{n/2} \left(\prod_1^n x_i^{-3/2}\right) \exp\left(-\lambda \sum_1^n \frac{(x_i - \mu)^2}{2\mu^2 x_i}\right)$$

and the maximum likelihood estimators of μ and λ are easily seen to be

$$\hat{\mu} = \bar{X}$$

and

$$\frac{1}{\hat{\lambda}} = n^{-1}\sum_{1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right).$$

These estimators were obtained by Schroedinger [1915].

The distribution of these estimators is also known. Tweedie [Tweedie, 1957a and 1957b] proved the following results.

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed as inverse gaussian with finite first and second moments. Then $\sum X_i$ and $\sum X_i^{-1} - n^2(\sum X_i)^{-1}$ are independently distributed.

Tweedie showed in his proof that $\bar{X} \sim IG(\mu, n\lambda)$. He defined

$$V = \sum_{1}^{n}\left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right) \tag{1.15}$$

and also proved

$$\lambda V \sim \chi^2_{n-1}$$

1.2.1.7 Related Distributions

Let

$$Y^2 = \frac{(X-\mu)^2}{X\mu^2}$$

Then

$$\lambda Y^2 \sim \chi^2_1$$

that is, λY^2 has the chi-square distribution with one degree of freedom. [Chhikara and Folks, 1989].

Let $X_1, X_2, \ldots, X_{n_X}$ and $Y_1, Y_2, \ldots, Y_{n_Y}$ be independent random samples from $IG(\mu, \lambda)$. Let

$$V_X = \sum_1^{n_X} \left(\frac{1}{X_i} - \frac{1}{\bar{X}} \right) \tag{1.16}$$

and let V_Y be similarly defined. Then it follows from its form as the ratio of two independent χ^2 random variables and is well known [Chhikara and Folks, 1989] that

$$R = \frac{(n_Y - 1)V_X}{(n_X - 1)V_Y} \sim F_{n_X - 1, n_Y - 1} \tag{1.17}$$

where F_{n_X-1,n_Y-1} is the standard F distribution with $n_X - 1$ and $n_Y - 1$ degrees of freedom.

1.2.1.8 Predictive distribution for non-informative prior

To obtain natural conjugate priors, it is advantageous to reparameterize the distribution. This poses no logical difficulty in the predictive framework, where the parameters will be integrated out in the process. The parameterization we use sets $\psi = 1/\mu$, and is due to Tweedie.

Lacking data, one would often prefer a non-informative prior distribution. Jeffrey's prior sets $p(\psi, \lambda) \propto \sqrt{|I(\psi, \lambda)|}$. Unfortunately, this prior produces a posterior which is not a proper distribution [Chhikara and Folks, 1989].

If one considers instead a diffuse prior for the parameters, the predictive distribution for the next observation given a series of observations from the inverse gaussian is known and due to Chhikara and Guttman [1982]. The prior used is

$$p(\psi, \lambda) \propto \lambda^{-1}.$$

Then let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ be n independent observations from $IG(\mu, \lambda)$, and Y be an additional observation taken independently of $\mathbf{x}$. For $y > 0$,

$$h(y|\mathbf{x}) = k \left[\frac{\bar{x}\hat{\lambda}}{(n\bar{x} + y)y^2} \right]^{1/2} \left[1 + \frac{(\bar{x} - y)^2 \hat{\lambda}}{\bar{x}y(n\bar{x} + y)} \right]^{-n/2}, \tag{1.18}$$

where

$$k = \frac{S_{t,n}\left((n+1)\sqrt{\frac{n}{z(n\bar{x}+y)}}\right)}{\beta\left(\frac{1}{2}, \frac{n-1}{2}\right) S_{t,n-1}\left(\sqrt{\frac{(n-1)}{v\bar{x}}}\right)}, \tag{1.19}$$

$S_{t,n}$ denotes the Student's t cumulative distribution function with n degrees of freedom, and

$$z = nv + \frac{n(\bar{x}-y)^2}{\bar{x}y(n\bar{x}+y)}$$

Note that the expression given in Equation 1.19 for k is correct, and rectifies an existing unnoticed error in both the *Technometrics* article by Chhikara and Guttman [1982] and the Chhikara and Folks monograph [1989].

As an example, consider the following 5 observations from an $IG(\mu, \lambda)$ distribution, with μ and λ unknown: $\{x_1 = 3, x_2 = 4, x_3 = 6, x_4 = 3.5, x_5 = 2.5\}$. Then the graph of the predictive density for the next observation can be seen in Figure 1.10.

Predictive limits for Y can be obtained by solving the appropriate integral equation using $h(y|\mathbf{x})$. To find the lower predictive limit, $l(\mathbf{x})$, given the data $\mathbf{x}$, for the next observation with probability $\alpha/2$, we solve numerically the integral equation:

$$\int_0^{l(\mathbf{x})} h(y|\mathbf{x})\, dy = \frac{\alpha}{2} \tag{1.20}$$

and similarly for the upper predictive limit, $u(\mathbf{x})$.

We derive the more general joint predictive distribution for the next m observations, given the first n, below. We also derive tighter predictive limits.

1.2.2 Modeling advantages over other skewed distributions

There are three major advantages to using the inverse gaussian distribution to model skewed data. The first is an appeal to the underlying physical properties of the process

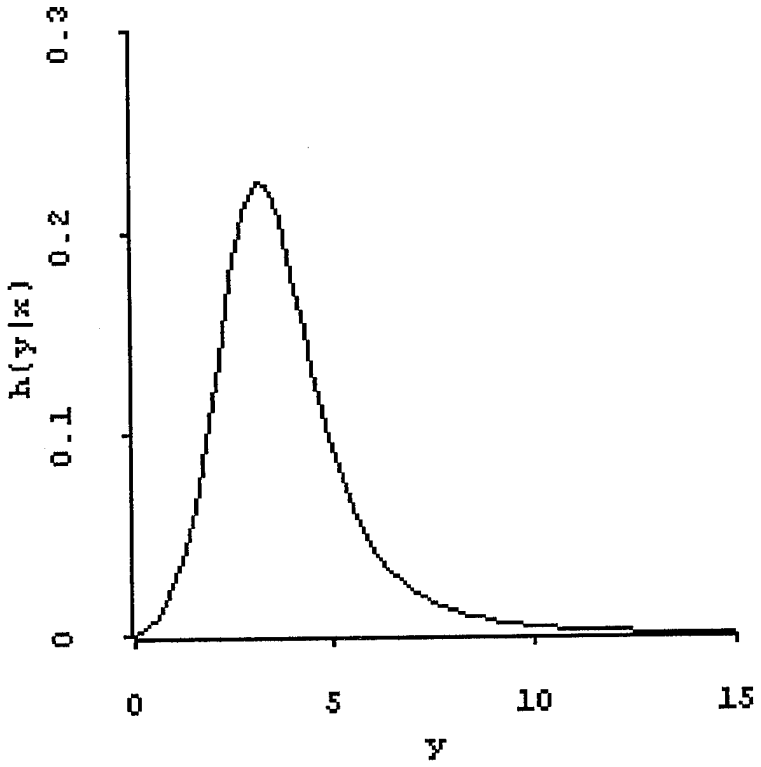


Figure 1.10: An example of the predictive density for the next observation from an $IG(\mu, \lambda)$, for five previous points: $\{x_1 = 3, x_2 = 4, x_3 = 6, x_4 = 3.5, x_5 = 2.5\}$.

being modeled. The second discusses the notion of failure rates, and their asymptotic behavior. The third is based on the tractability of the sampling distribution of the inverse gaussian. We discuss each of the three in turn.

If the underlying process can be thought of as a Wiener process, then the use of the inverse gaussian seems especially appropriate. The time to failure of a complex system may depend on the accumulated additive effects of many small perturbations. Compare this with the log-normal distribution, whose application to modeling depends on multiplicative effects, which are often difficult to defend from first principles.

Secondly, consider the failure rate, $r(t)$ of a system as a function of time. We define

$$r(t) = \frac{f(t)}{1 - F(t)},$$

where $r(t)$ denotes the instantaneous rate of failure for a process conditional on its having lasted a certain time. For a Poisson process with time to failure modeled by the exponential distribution with parameter λ, $r(t) = \lambda$; a constant failure rate.

The assumption of constant failure rate is rather strong. Some applications call for a monotonic failure rate: some with an increasing failure rate (IFR); others for a decreasing failure rate (DFR). For these, it is possible to use the Weibull distribution, with density $f(t, \alpha, \beta) = \alpha\beta^{-\alpha}exp(-(x/\beta)^{\alpha})$, $(x, \alpha, \beta > 0)$. Then $r(t) = \alpha\beta^{-\alpha}x^{\alpha-1}$, and is decreasing for $\alpha < 1$ and increasing for $\alpha > 1$.

In many situations which are characterized by a "burn in" process, it seems appropriate to have an initially increasing then decreasing failure rate [Chhikara and Folks, 1977].

An initially IFR then DFR process is often modeled by the log-normal distribution. For such a process,

$$r(t) = f_{\mu,\sigma}(t)/(1 - \Phi((t - \mu)/\sigma)).$$

This failure rate is non-monotonic; increasing then decreasing asymptotically to zero. For many reliability situations, this asymptotic failure rate of zero seems illogical.

An alternate model uses the inverse gaussian process. Its failure rate is given by:

$$r(t) = \frac{f(t)}{1 - F(t)} \tag{1.21}$$

$$= \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left(\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right)}{\Phi\left(\sqrt{\frac{\lambda}{t}}\left(1 - \frac{t}{\mu}\right)\right) - e^{2\lambda/\mu}\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(1 + \frac{t}{\mu}\right)\right)} \tag{1.22}$$

This failure rate is also non-monotonic, initially increasing then decreasing. However, its asymptotic failure rate is given by

$$r(t) \rightarrow \frac{\lambda}{2\mu^2}$$

It is easily shown that the maximum value of $r(t)$ occurs at the value t^* which is the solution to the equation

$$r(t) = \frac{\lambda}{2\mu^2} + \frac{3}{2t} - \frac{\lambda}{2t^2}$$

and the maximum failure rate is $r(t^*)$.

This provides a strong argument for using the inverse gaussian over the log-normal distribution to model lifetimes. It is hard to conceive of physical processes where the failure rate would decrease to zero.

The third argument for using the inverse gaussian is that the sampling distribution of the MLEs of the parameters are known and easy to work with, as above. Using the inverse gaussian avoids the need to transform the data prior to finding MLEs, as is the case with the log-normal distribution.

1.3 Scope of this thesis

This thesis extends and corrects the work of Edgeman, who first developed Shewhart control charts for the inverse gaussian distribution. We develop self-starting and

predictive Shewhart charts, and discuss their features. We then develop CUSUM charts for the inverse gaussian distribution, along with several variants, and discuss their properties. We also propose and develop a new type of bi-variate Shewhart chart, and apply it to the inverse gaussian case.

We apply these methods in two settings. First, we model the loss exchange ratio (LER) for a military mission as an inverse gaussian random variable. The LER is the ratio of enemy casualties to friendly casualties. We use standard military modeling packages to generate this data using high-resolution simulation – simulation in the war-gaming, not statistical, sense. We verify that the inverse gaussian distribution provides a reasonable fit. We then show how statistical process control, and in particular the methods of this thesis, can alert the decision maker to a shift in the distribution, and the significance of such an alert. We explain the significance of this application.

Second, we consider an example from the literature where the time to complete a task on a General Motors assembly line has been well modeled by the inverse gaussian distribution. We apply our methods to that case, and discuss the significance of that application.

Chapter 2

Shewhart Control Charts for IG Processes

In this chapter, we discuss Shewhart Control charts for processes modeled by the inverse gaussian distribution.

2.1 Edgeman's work

Edgeman [1989] proposed control charts for the inverse gaussian distribution. He proposed charting transformations of the sufficient statistics $\bar{X}$ and V, discussed above. The transformations were originally given by Chhikara and Folks [1976].

Let $X \sim IG(\mu, \lambda)$. Assume λ is fixed and known. Chhikara and Folks [1976] showed that the uniformly most powerful unbiased test for $H_o : \mu = \mu_o$ against $H_A : \mu \neq \mu_o$ is of the form $\bar{X} > k_1$ or $\bar{X} < k_2$, since the IG has a monotone likelihood ratio, and is in the one-parameter exponential family when λ is known. To obtain a pivotal, they proposed the transformation to the statistic

$$Y = \frac{(N\lambda\ )^{1/2}(\bar{X} - \mu_o)}{\mu_o\sqrt{\bar{X}}} \tag{2.1}$$

The cumulative distribution function of Y is given by

$$G(y) = \Phi(y) + \exp(2\lambda/\mu)\Phi\left(-\sqrt{y^2 + \frac{4\lambda}{\mu}}\right)$$

where $\Phi(y)$ is the standard normal CDF. It follows from the CDF of Y that $|Y|$ has the folded normal distribution. The rejection region for the test is given by

$$|Y| > z_{1-\alpha/2}$$

where $z_{1-\alpha/2}$ is the α critical value from the standard normal distribution.

Similarly, when λ is unknown, the test statistic becomes

$$W = \frac{\sqrt{n\,(n-1)}\left(\overline{X} - \mu_o\right)}{\mu_o\sqrt{\overline{X}\,V}} \tag{2.2}$$

where V is defined as in Equation 1.15. The critical region is

$$|W| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the critical value from the t distribution with $n-1$ degrees of freedom.

The tests for one-sided alternative hypotheses are not quite as simple in form, but exist and are discussed in [Chhikara and Folks, 1989].

Edgeman replaced μ_o with μ and constructed confidence intervals for μ of the form

$$\left(\frac{\overline{X}}{1 + Z_{1-\alpha/2}\sqrt{\overline{X}/\left(N\lambda\right)}}, \frac{\overline{X}}{\max\left(0, 1 - Z_{1-\alpha/2}\sqrt{\overline{X}/\left(N\lambda\right)}\right)}\right)$$

when λ is known and of the form

$$\left(\frac{\overline{X}}{1 + t_{1-\alpha/2}\sqrt{\frac{\overline{X}V}{N(N-1)}}}, \frac{\overline{X}}{\max\left(0, 1 - t_{1-\alpha/2}\sqrt{\frac{\overline{X}V}{N(N-1)}}\right)}\right)$$

when λ is unknown.

The right hand limits contain the "max" function in the denominator to restrict $\mu > 0$.

Edgeman then assumed that there was historical data based on about "$M = 20$ to $M = 25$ samples". From that data, he substituted the grand average of all observations, $\overline{\overline{X}}$ for $\overline{X}$, and $\overline{V}$ for V, to obtain lower and upper control limits (LCL

and UCL, respectively) for the process centrality when λ is known and unknown.

$$UCL_{\lambda known} = \frac{\overline{\overline{X}}}{\max\left[0, 1 - z_{1-\alpha/2}\sqrt{\frac{\overline{\overline{X}}}{N\lambda}}\right]} \tag{2.3}$$

$$LCL_{\lambda known} = \frac{\overline{\overline{X}}}{1 + z_{1-\alpha/2}\sqrt{\frac{\overline{\overline{X}}}{N\lambda}}} \tag{2.4}$$

$$UCL_{\lambda unknown} = \frac{\overline{\overline{X}}}{\max\left[0, 1 - t_{1-\alpha/2}\sqrt{\frac{\overline{\overline{X}}\,\overline{V}}{N}}\right]} \tag{2.5}$$

$$LCL_{\lambda unknown} = \frac{\overline{\overline{X}}}{1 + t_{1-\alpha/2}\sqrt{\frac{\overline{\overline{X}}\,\overline{V}}{N}}} \tag{2.6}$$

The center line for the chart (CL) for the mean is given by $\overline{\overline{X}}$. The subsequent sample averages are plotted against these limits, and the process is signaled as "out-of-control" for process centrality if/when a sample average exceeds the control limits.

We shall see in a later section that the performance of these charts is very poor. In checking the original article, we find no performance data for Edgeman's scheme, only for a competitor.

We conjecture that Edgeman's original article contains an error. We find that if we proceed as Edgeman did, but substitute the historical mean for μ_0 into Equation 2.1, and then construct upper and lower bounds on $\bar{X}$, we obtain:

$$LCL_{\lambda known} = \frac{\overline{\overline{X}}\left(\sqrt{\overline{\overline{X}}z_{\alpha/2}^2 + 4\lambda n} + \sqrt{\overline{\overline{X}}}z_{\alpha/2}\right)^2}{4n\lambda} \tag{2.7}$$

$$UCL_{\lambda known} = \frac{\overline{\overline{X}}\left(\sqrt{\overline{\overline{X}}z_{1-\alpha/2}^2 + 4\lambda n} + \sqrt{\overline{\overline{X}}}z_{1-\alpha/2}\right)^2}{4n\lambda} \tag{2.8}$$

$$\tag{2.9}$$

Similarly, if λ is unknown, we proceed from Equation 2.2 and obtain:

$$LCL_{\lambda unknown} = \frac{\overline{\overline{X}}V_H\left(\sqrt{\frac{\overline{\overline{X}}V_H t^2_{\alpha/2,n-1}+4n^2-4n}{V_H}}+\sqrt{\overline{\overline{X}}}t_{\alpha/2,n-1}\right)^2}{4n(M-1)} \quad (2.10)$$

$$UCL_{\lambda unknown} = \frac{\overline{\overline{X}}V_H\left(\sqrt{\frac{\overline{\overline{X}}V_H t^2_{1-\alpha/2,n-1}+4n^2-4n}{V_H}}+\sqrt{\overline{\overline{X}}}t_{1-\alpha/2,n-1}\right)^2}{4n(M-1)} \quad (2.11)$$

We use $V_H = \sum_{i=1}^{M}\left(\frac{1}{X_i}-\frac{1}{\overline{\overline{X}}}\right)$, which is our unbiased estimator for $(M-1)/\lambda$ using the M historical data points. These control limits perform as expected, and should be used in lieu of Edgeman's original ones. Note that n refers to the size of the rational subgroup.

By illustration, if one assumes an $IG(3,5)$ distribution, samples of size 5, $\alpha = .01$, and uses the uncorrected scheme of Equations 2.4 and 2.5, one obtains $LCL = 1.58538$ and $UCL = 27.8508$. In control, the ARL for the uncorrected scheme is 24.000, instead of the desired ARL of 100. Clearly, something is amiss.

If we use our corrected scheme, we obtain $LCL = 1.26308$ and $UCL = 7.12542$, and the ARL is 99.9891. This is the desired performance.

Edgeman controlled for λ in a similar fashion, using the fact that the distribution of $V \sim (1/\lambda)\chi^2_{n-1}$. He obtained

$$UCL = \frac{\overline{V}\chi^2_{\alpha/2,N-1}}{(N-1)} \quad (2.12)$$

$$LCL = \frac{\overline{V}\chi^2_{1-\alpha/2,N-1}}{(N-1)} \quad (2.13)$$

$$CL = \overline{V} \quad (2.14)$$

We will see in our example in Chapter 8 that this control chart scheme behaves unexpectedly, in that the ARL for out-of-control states obtained by small shifts towards the heavy tail is actually greater than the ARL when in-control. This will support our argument to use other techniques which do not suffer from this defect.

Edgeman compared his uncorrected charts with traditional $\overline{X}$ and R charts for normal processes, and found that the normal charts did not perform well. As noted above, he provided no evidence that his uncorrected design performed well, either.

Edgeman also noted that the "UCLs given ... can become infinite. In the event that an infinite UCL occurs, it may be desirable to construct control charts for reciprocal process centrality."

2.2 Improvements

Additional improvements to Edgeman's work will focus first on simplifying the work. Next, we extend it to the case where the parameters are not known *a priori*, and we design a "self-starting" control scheme which continually adapts to the data. An alternative, predictive control charts, will be developed in the next chapter.

2.3 Simplification

The transformation of the sample data into Y and W statistics was done to obtain pivotal quantities; that is, quantities whose distribution did not depend on the parameter. While this is useful in estimation, it is less so in significance testing, particularly where we assume *a priori* that we know the distribution of the process. Additionally, quality control methods should strive, where convenient, for simplicity in their implementation, as the calculations may be performed manually by the shop worker.

With this in mind, let us revisit the $\overline{X}$ charts. For λ known, we know that the uniformly most powerful (UMP) unbiased test for $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_o$ is of the form $\overline{X} > k_1$ or $\overline{X} < k_2$. In control, the distribution of $\overline{X}$ is known to be $IG(\mu, n\ \lambda)$. Accordingly, it is simpler to directly determine k_1 and k_2 to meet some criteria. Once the criteria are determined (discussed below) it is much easier

to implement the control charts. Additionally, the problem of infinite or negative control limits is also avoided.

k_1 and k_2 are set so that the probability that the process, in control, exceeds them is $\alpha = \frac{1}{ARL}$. Traditionally, they have been set symmetrically, satisfying $P(\overline{X} > k_1) = \alpha/2$ and $P(\overline{X} < k_2) = \alpha/2$.

The UMP unbiased test sets the control limits so that $\int_{k1}^{k2} f(x;\mu,\lambda)dx = 1/ARL$ and $\int_{k1}^{k2} x\ f(x;\mu,\lambda)dx = E(X)$. Economic arguments based on non-uniform loss might mitigate a different strategy.

The symmetric probability and UMP-unbiased control limits are not the tightest possible limits for a skewed distribution. For a skewed distribution, the highest probability density region (HPD) provides the shortest interval. See Figure 2.1.

The HPD is of the form $R(\alpha) = \{x : f(x) > k(\alpha)\}$. $k(\alpha)$ is found numerically, where α is the desired probability of an out–of–control signal when the process is in–control.

A routine for computing the HPD control limits for the inverse gaussian is available from the author.

In summary, we chart $\overline{X}$ to control for central tendency. Since the distribution of $\overline{X}$ is known to be $IG(\mu, n\lambda)$, we set the control limits (R) by finding either symmetric limits, UMP-unbiased limits, or the HPD region, $R(\alpha)$ for the $f_{\overline{X}}(\overline{x})$. The process is deemed to be in control as long as the sample average is within R.

This approach has two distinct advantages over Edgeman's work. First, we avoid the possibility of infinite upper or lower control limits. Second, with the HPD region we obtain the tightest (in the sense of shortest possible) control intervals among all intervals with the same α. We compare the performance of all these methods later in the chapter.

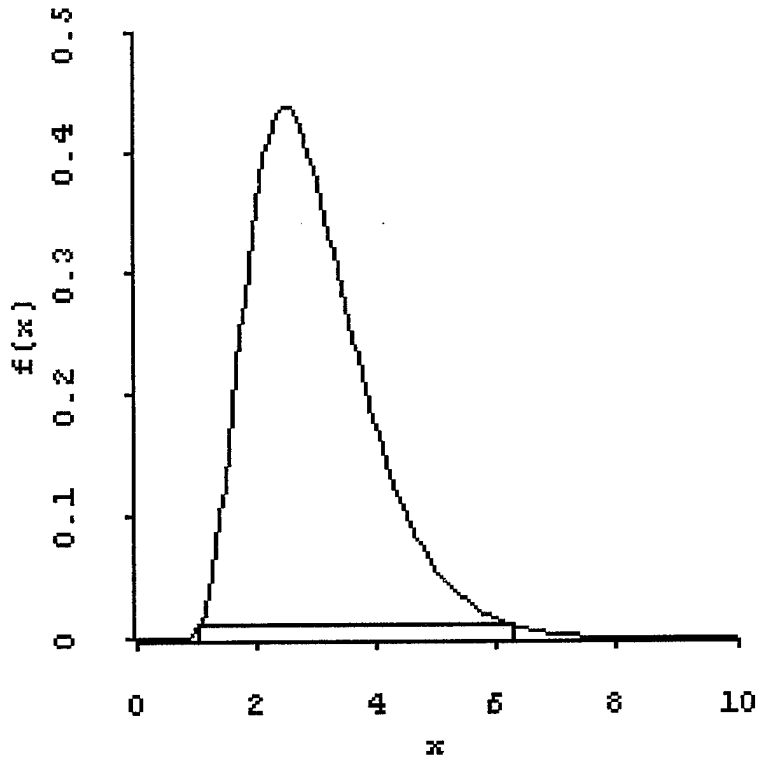


Figure 2.1: Illustration of the HPD region for a skewed distribution. Here $\alpha = .01$, and $X \sim IG(3, 25)$, corresponding to the distribution of the mean of a sample of size five from an $IG(3, 25)$. The upper and lower limits of the HPD region were found numerically using *Mathematica* to be $x = 1.04805$ and $x = 6.26794$. The value of $f(x)$ at the endpoints of the HPD is $f = .0119252$.

2.3.1 Average Run Lengths

We illustrate the comparative advantages of the known parameter approaches above. We examine the ARLs found for several $IG(\mu, \lambda)$ distributions, in-control and then with departures from control at various points in the process lifetime. The exact value of the out-of-control ARLs can be found by integration for all of the known parameter approaches.

It is apparent from Tables 2.3.1 and 2.3.1 that, among the four alternatives, the HPD test is the most powerful for detecting an increase in the mean, but performs poorly for detecting decreases. For a decrease in the mean, the ARL exceeds 100 for some of the test cases.

The symmetric limits appear to perform reasonably well for detecting both decreases and increases.

The corrected Edgeman test is much slower detecting an upward shift, compared to the symmetric and HPD tests. However, it does not appear to suffer loss of performance for downward shifts. Of course, we expect good performance from the corrected Edgeman test, as it is based on the uniform most powerful unbiased test for detecting a mean shift, given λ is unknown.

We note that in many applications (and both of the examples which follow later in this work), one is more interested in detecting increases in μ than in detecting decreases. This favors the HPD chart over the corrected Edgeman chart.

2.4 Self-starting Shewhart Charts

In this section, we propose charts based on the running estimates for μ and λ. Following Hawkins [1987], we call these "self-starting" charts.

In the preceding section, we assumed the process characteristics were known

Comparing ARLs for Mean Tests					
μ	λ	Edgeman	Corrected Edgeman	HPD	Symmetric
3	5	24.00	99.98	100	100
4	5	81.91	20.79	11.18	15.18
5	5	191.94	6.56	4.31	5.30
6	5	344.117	3.62	2.69	3.11
7	5	378.658	2.58	2.06	2.30
1	5	1.01	1.12	1.58	1.17

Table 2.1: Comparison of the Edgeman, Symmetric, and HPD schemes for a mean shift, λ known, $\alpha = .01$, $n = 5$, with the in-control distribution as $IG(3, 5)$. Note the poor performance of the uncorrected Edgeman scheme. Also note that the HPD is not as quick to detect a downward mean shift.

Comparing ARLs for Mean Tests					
μ	λ	Edgeman's	Corrected Edgeman	HPD	Symmetric
42	66	23.17	100	100	100
45	66	31.00	86.75	53.8326	69.87
50	66	47.83	46.86	23.4519	33.17
60	66	97.08	14.49	8.1826	10.81
70	66	168.04	6.96	4.5026	5.56
40	66	18.82	93.80	154.278	109.93
35	66	10.61	53.79	267.764	75.22

Table 2.2: Comparison of the Edgeman, Symmetric, and HPD schemes for a mean shift, λ known, $\alpha = .01$, $n = 5$, with the in-control distribution as $IG(42, 66)$. This distribution is discussed later in the General Motors example. Again, note the poor performance of the uncorrected Edgeman scheme and the one-sided performance of the HPD test.

exactly. This, of course, is not always the case. Edgeman addressed this issue by setting $\mu = \overline{\overline{X}}$ and $\lambda = 1/\overline{V}$. These are only estimates, and subject to sampling error. If there is extensive historical data, then the error will be small. However, Edgeman states

> The values of $\overline{\overline{X}}$ and $\overline{V}$ should be based on the results of about $M = 20$ to $M = 25$ samples

citing Montgomery [1985]. A check of the revised referenced work indicates that, in the section discussing the statistical basis of the chart for the normal distribution, Montgomery [1991, p. 203] states

> These estimates should usually be based on *at least* 20 to 25 samples. (Emphasis added.)

The use of the rule of thumb of "about" 20 to 25 historical samples for setting control limits for a non-normal, potentially heavy tailed distribution such as the inverse gaussian does not seem supported by this reference. For CUSUM charts, a different context, Hawkins [1987] has found evidence for normal processes that "25 start-up observations (as seems to be the standard practice) is too short a learning set, particularly as regards the process standard deviation."

Additionally, the use of a fixed learning set to set process limits ignores the improved precision one may get from incorporating subsequent data into the estimates of the process parameters. Further, such a rule is of no help in the start–up phase of the process, when we wish to control the process while also gathering the process initial data.

To address these issues, we may use self–starting control charts. Hawkins [1987] has proposed self-starting CUSUM control charts. He argues for them as follows:

> Thus the self-starting cusums produce superior performance to those obtained with some 25 special start-up values while not involving any measurements beyond those produced anyway. ... We would therefore argue that self-start CUSUMs should always be used in preference to plugging the mean and observations of a start-up sample into the formulae assuming known parameters.

We argue similarly for self-starting Shewhart charts, while noting that the Shewhart-type chart is more robust to a model mis-specification of parameter than the CUSUM. Hence the case for the self-starting CUSUM may be more compelling than that for the self-starting Shewhart chart.

We design two self-starting Shewhart charts, one for location and one for shape. The self-starting Shewhart chart is based on the UMP-unbiased test [Chhikara, 1975] for the equality of two inverse gaussian population means. The self-starting Shewhart chart for shape is based on the likelihood ratio test for the equality of the scale parameters λ and τ of two inverse gaussian populations $IG(\mu, \lambda)$ and $IG(\nu, \tau)$. These are discussed in detail below.

2.4.1 Self-Starting Shewhart charts for location for the $IG(\mu, \lambda)$

Assume we have two inverse gaussian processes, X and Y, with common but unknown scale parameter λ. Chhikara [1975] derived the uniformly most powerful unbiased (UMP-unbiased) test for the equality of the two process means. We test the hypothesis

$$H_0 : \mu = \nu \text{ versus } H_1 : \mu \neq \nu$$

We draw a sample from each population. The rejection region is given by

$$|T| = \left| \frac{\sqrt{n_1 n_2 (n_1 + n_2 - 2)}\ (\overline{X} - \overline{Y})}{\sqrt{\overline{X}\,\overline{Y}(n_1\overline{X} + n_2\overline{Y})(V_1 + V_2)}} \right| > t_{1-\alpha/2, n_1+n_2-2} \tag{2.15}$$

$|T|$ has the folded t distribution with $n_1 + n_2 - 2$ degrees of freedom, a result reported in Chhikara and Folks [1989]. Note that the expression in Equation 6.31 of Chhikara and Folks [1989] is incorrect; the expression in Equation 2.15 is correct.

To construct a self-starting scheme, we collect a first sample of size n, call it X_1. We then collect a second sample of size n, Y_1, compute T_1 and test for the equality of means. If the null hypothesis is not rejected, we merge sample X_1 and Y_1 into the new reference sample, X_2, of size $2n$. We then collect our third sample of size n, call it Y_2, compute T_2 and test for the equality of means for X_2 and Y_2. If it is not rejected, we merge Y_2 with X_2, and draw another sample. And so forth.

The scheme suffers from the disadvantage that the control limits depend on a t statistic, whose value changes as the sample sizes change. This problem is reduced as the number of samples grows, because $t \rightarrow N(0,1)$.

There is a further transformation of the test statistic that can allow for constant control limits.

For a given observation of T, we compute its p-value using the CDF for T: $P = F_T(T)$. It is well known that $P \sim U(0,1)$. We then convert that percentile to a $N(0,1)$ variate, using the inverse CDF for the standard normal distribution: $Z = \Phi^{-1}(P)$. We are now in the very familiar setting of charting standard normal variates,

$$Z = \Phi^{-1}(F_T(T)) \sim N(0,1)$$

While these transformations complicate the computations for the operator doing his charts by hand, for those using a computer this scheme has the advantage of constant limits, a distribution unaffected by the number of samples to date, and a very familiar context. This allows simpler charting and easier probability calculations.

We obtain control limits in the usual manner, first specifying the ARL we wish. We then set $\alpha = \frac{1}{\text{ARL}}$ and find $z_{\alpha/2}$ and $z_{1-\alpha/2}$, the usual critical values. Since in-control we are charting normal variates, all standard control chart theory applies.

If the mean of the process shifts, we see that the numerator of T no longer has expected value of 0. Accordingly, the ARL out-of-control is reduced. Calculation of the exact out-of-control ARL is cumbersome due to the mixed in-control and out-of-control distributions in the denominator of T and is further complicated by the transformations. For approximating those out-of-control ARL calculations, we suggest simulation. Simulation also provides a characterization of the out-of-control run length distribution itself.

As an alternative, one could forego the above transformations and chart the corresponding p-values for the T_i, and signal if the p-values fell in some rejection region. As the p-values are uniformly distributed while the process is in-control, it is easy to derive a control scheme for them. We prefer to chart the corresponding normal variates, because they provide higher visual resolution in the tails of the distribution.

2.4.1.1 Example

We provide the following example of a self-starting chart for the process mean. The in-control distribution is $IG(3, 5)$. We take samples of size one by simulation. We desire an ARL of 100, so we signal if $|z| > 2.5758$ or if $.005 < p < .995$.

Our first case has an out-of-control point deliberately inserted at observation 6. Table 2.3 provides the data. The control chart is presented in Figure 2.2. We see that the scheme does not catch an early outlier, which is to be expected. We also see that the next several observations report low p-values and z-scores, which is also to be expected until the effect of the outlier is averaged out.

Our second case has the outlier at observation 16. Table 2.4 provides the data.

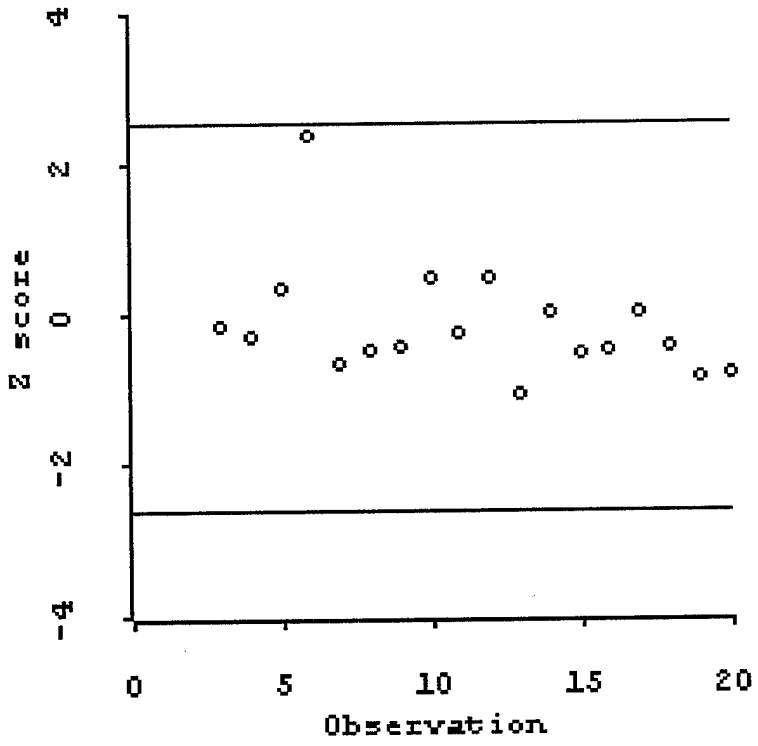


Figure 2.2: Control chart for the self-starting example with early outlier. $\alpha = .01$.

n	μ	λ	X	P-value	Signal?
1	3	25	3.5156	n/a	
2			1.4671	n/a	
3			2.2176	.4520	
4			2.0622	.3965	
5			2.7420	.6437	
6	7	40	13.0000	.9916	NO
7	3	25	2.2903	.2662	
8			2.6414	.3282	
9			2.6564	.3343	
10			5.2536	.6860	
11			3.1756	.4041	
12			5.2532	.6888	
13			1.9516	.1504	
14			3.7829	.5119	
15			2.7475	.3141	
16			2.7620	.3208	
17			3.6853	.5170	
18			2.8630	.3408	
19			2.2799	.2044	
20			2.3533	.2279	

Table 2.3: Self-starting results for a data stream with early outlier. $\alpha = .01$.

The control chart is presented in Figure 2.3. We see that the scheme does catch this outlier when it occurs later in the process. We also see the effect on the observations following the outlier is less dramatic.

Self-starting charts offer us the opportunity to detect outliers in our initial phase. This provides control not available otherwise. Additionally, once we correct the conditions which caused them, we can delete those outliers from our process historical data, obtaining better estimates for subsequent control.

We return to this data set when we analyze the performance of the predictive charts developed in the next chapter.

2.4.2 Self-Starting Control charts for shape for the $IG(\mu, \lambda)$

Assume we have two inverse gaussian processes, X and Y, with common but unknown mean, μ. Chhikara [1975] derived the uniform most powerful unbiased (UMP-unbiased) test for the equality of the two process shape parameters, λ. We test the

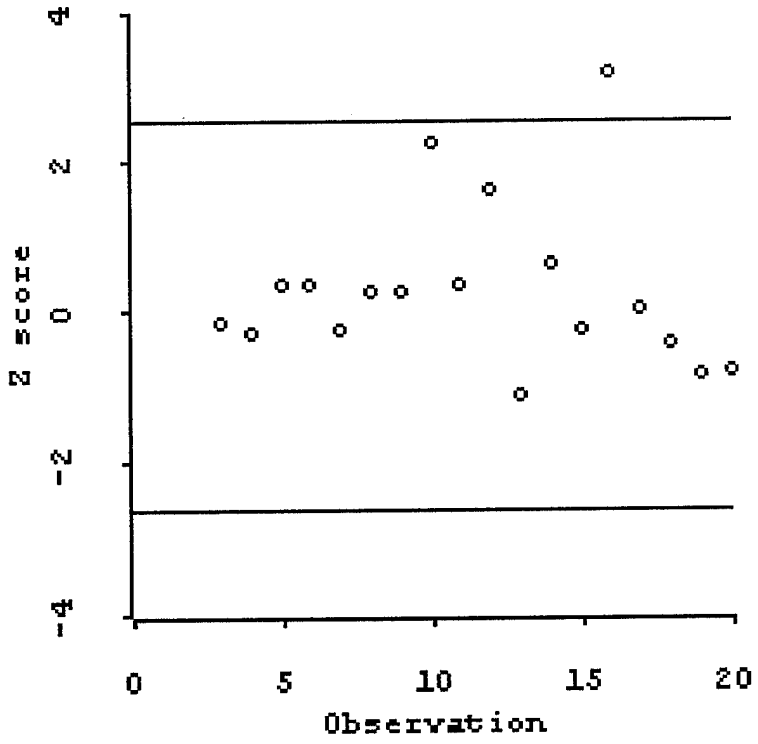


Figure 2.3: Control chart for the self-starting example with later outlier. $\alpha = .01$.

n	μ	λ	X	P-value	Signal?
1	3	25	3.5156	n/a	
2			1.4671	n/a	
3			2.2176	.4520	
4			2.0622	.3965	
5			2.7420	.6437	
6			2.7620	.6385	
7			2.2903	.4179	
8			2.6414	.6013	
9			2.6564	.6033	
10			5.2536	.9882	
11			3.1756	.6451	
12			5.2532	.9472	
13			1.9516	.1427	
14			3.7829	.7382	
15			2.7475	.4165	
16	7	40	13.0000	.9993	YES
17			3.6853	.5170	
18			2.8630	.3408	
19			2.2799	.2044	
20			2.3533	.2279	

Table 2.4: Self-starting results for a data stream with later outlier. $\alpha = .01$.

hypothesis:

$$H_o : \lambda_X = \lambda_Y \text{ versus } H_A : \lambda_X \neq \lambda_Y$$

Again, we draw a sample from each population. Let

$$V_X = \sum_{1}^{n_X} \left(\frac{1}{X_i} - \frac{1}{\overline{X}} \right)$$

and V_Y be similarly defined. Then we recall from Equation 1.17 that the distribution under the null hypothesis of

$$R = \frac{(n_Y - 1)V_X}{(n_X - 1)V_Y}$$

is known to be F_{n_X-1,n_Y-1}.

We can use this known distributional result to construct a simple self-starting Shewhart scheme to control for the shape parameter.

We set our significance level, α, *a priori*. We then draw our first sample from our process, and call it Y_1. We draw our second sample, and call it X. We compute R for the two samples, X and Y_1. We find the corresponding two-sided p-value from

a standard F table or numerical routine with the appropriate degrees of freedom, and chart it. If $p \notin [\alpha/2, 1-\alpha/2]$, we signal. Else, we define $Y_2 = Y_1 + X$, and draw a new sample , X.

In control, our ARL is $1/\alpha$.

Now, assume that the distribution shifts in the shape parameter to λ_a. For the first observation out-of-control, we have

$$X \sim IG(\mu, \lambda_a)$$

Then the distribution of V_X shifts as well:

$$V_X \sim \frac{\chi^2_{n_X-1}}{\lambda_a}$$

Then

$$R = \frac{\lambda(n_{Y_i}-1)V_X}{\lambda(n_X-1)V_Y} = \frac{\lambda}{\lambda_a} \times R_a$$

where R_a has the F distribution with the appropriate degrees of freedom. In other words, there is a scale shift in R. For this first observation, we can compute explicitly the probability of a signal:

$$P\left(\frac{\lambda}{\lambda_a} \times R_a \notin [C_{lower}, C_{upper}]\right) = P\left(R_a \notin \left[\frac{\lambda_a C_{lower}}{\lambda}, \frac{\lambda_a C_{upper}}{\lambda}\right]\right)$$

For subsequent observations out-of-control, the distribution of R is not as clean, due to the mixing of differently scaled χ^2 variables in the expression for Y_i.

2.5 Conclusion

In this chapter, we have improved and extended the work by Edgeman, who first described Shewhart charts for the inverse gaussian distribution. We have corrected the test for the mean of the process. We have explored both symmetric and HPD control limits, and contrasted them to the corrected Edgeman scheme. We have also

described self-starting Shewhart charts to allow both process control in the startup phase and improved control during the run phase. The examples in the preceding section illustrated the comparative utility of these approaches.

Chapter 3

Predictive control charts for the IG

Instead of using the self-starting methodology of the previous chapter, it is possible to consider the problem anew from a Bayesian perspective. This offers the possibility of incorporating pre-existing information about the process into our control schemes, while simultaneously recognizing that there is uncertainty associated with the pre-existing information.

We consider two classifications of schemes in this chapter: rational groups of size one and then larger rational groups. We derive the predictive limits for the next observation or next group of observations.

The predictive framework has the advantage of dealing with observables. While traditional quality control methods control for the value of a parameter, the predictive approach controls more generally for model departures. If the model is correct, and unchanged, we may obtain an interval with a given probability for the next observation. If the next observation is within that interval, we update and continue to sample. If the next observation is outside that interval, we stop the process and check for model departures in the underlying physical process.

Since we also obtain the distribution of the next observable, we are able to make expected loss calculations for any loss function. We can then control for expected loss as a function of the next observable, instead of controlling for the value of the observable, or even for a parameter shift. The ability to introduce loss functions in this manner also argues strongly for the predictive approach.

3.1 Sample of size one

In this section, we utilize Equations 1.18 and 1.19 to construct a control scheme for the next observation from an inverse gaussian process. In this approach, we assumed an appropriate non-informative prior distribution, as was used in Chapter 1 where these results were reviewed.

Instead of a non-informative prior distribution, it is possible to use the natural conjugate prior for the inverse gaussian as the prior. Those methods are not developed further in this thesis. Such an approach would have the advantage of greater predictive power early in the process startup phase, reflecting a more informed prior opinion.

Recall that the predictive distribution for the next observation Y from an inverse gaussian process with previous observations $X_1, X_2, \ldots, X_n$ is given by:

$$h(y|\mathbf{x}) = k\left[\frac{\bar{x}\hat{\lambda}}{(n\bar{x}+y)y^3}\right]^{1/2}\left[1+\frac{(\bar{x}-y)^2\hat{\lambda}}{\bar{x}y(n\bar{x}+y)}\right]^{-n/2}, \tag{3.1}$$

where

$$k = \frac{S_{t,n}\left((n+1)\sqrt{\frac{n}{z(n\bar{x}+y)}}\right)}{\beta\left(\frac{1}{2},\frac{n-1}{2}\right)S_{t,n-1}\left(\sqrt{\frac{(n-1)}{v\bar{x}}}\right)}, \tag{3.2}$$

$S_{t,n}$ denotes the Student's t distribution with n degrees of freedom, $\bar{x}$ is the arithmetic mean, and $v = \sum(1/x_i) - 1/\bar{x}$, and

$$z = nv + \frac{n(\bar{x}-y)^2}{\bar{x}y(n\bar{x}+y)}$$

As in the previous chapter, there are two approaches to finding limits for the next observation. The first uses symmetric quantiles, and requires us to compute quantiles for predictive distribution. A numerical routine which computes these quantiles is available from the author. The second approach uses the highest probability density region. It, too, has a numerical routine available from the author.

From each of these schemes (controlling with probability limits and controlling with highest probability density limits), we have obtained a region R for the next observation Y. If $Y \in R$, we continue to sample. If $Y \notin R$, then we stop the process and check for a physical reason for the model departure.

For the first approach, notice that we do not actually have to compute the region, R, which is an extensive exercise in numerical computing. Rather, we compute the p value associated with the observation, and compute and chart it. If $P(Y < y)$ is too low or too high, we stop the process as out-of-control.

For the second approach, we find the $P(h(Y) < h(y))$. Again we chart this probability. If this probability is too low (one sided test here), we stop the process as out-of-control. Computing $P(h(Y) < h(y))$ is easier than finding the corresponding rejection region.

Charting p values has the additional benefit of being easy to explain to the practitioner. We are charting how unlikely the subsequent data is, given all of the preceding observations.

3.1.1 An example

Let us assume that we have a process with true but unknown parameters $\mu = 3$, and $\lambda = 25$. Let us further assume that we have an out-of-control state which has $\mu = 7$ and $\lambda = 40$. We will run the process in control for various lengths, and then see if the chart detects the departures. We will set $\alpha = .01$, resulting in an ARL of 100 in control. Notice the scheme does not know what the initial parameters are. Tables 3.1.1 and 3.2 record what happens.

In constructing the samples, we generated 19 observations from an $IG(3,\ 25)$, and one outlier. The outlier was chosen so as not to signal early in the process, but to signal late in the process. We see that the vagueness in the prior results in a

n	μ	λ	X	P-value	Signal?
1	3	25	3.5156	n/a	
2			1.4671	n/a	
3			2.2176	.4704	
4			2.0622	.4718	
5			2.7420	.7141	
6	7	40	13.0000	.9947	NO
7	3	25	2.2903	.3796	
8			2.6414	.4529	
9			2.6564	.4554	
10			5.2536	.7978	
11			3.1756	.5253	
12			5.2532	.7938	
13			1.9516	.2174	
14			3.7829	.6277	
15			2.7475	.4131	
16			2.7620	.4175	
17			3.6853	.6219	
18			2.8630	.4344	
19			2.2799	.2741	
20			2.3533	.3010	

Table 3.1: Representative predictive results for early outlier. We signal if $p \notin (.005, .995)$.

n	μ	λ	X	P-value	Signal?
1	3	25	3.5156	n/a	
2			1.4671	n/a	
3			2.2176	.4704	
4			2.0622	.4718	
5			2.7420	.7141	
6			2.7620	.7048	
7			2.2903	.4816	
8			2.6414	.6578	
9			2.6564	.6557	
10			5.2536	.9926	
11			3.1756	.7099	
12			5.2532	.9657	
13			1.9516	.1851	
14			3.7829	.7988	
15			2.7475	.4881	
16	7	40	13.0000	.99973	YES
17			3.6853	.6219	
18			2.8630	.4344	
19			2.2799	.2741	
20			2.3533	.3010	

Table 3.2: Representative predictive results for later outlier. We signal if $p \notin (.005, .995)$.

wide posterior distribution for the next observable early in the process. That is why observation 6 in Table 3.1.1 does not signal, but observation 16 in Table 3.2 does.

We can introduce a loss function into this problem. Assume that there is a loss function $L(Y)$ associated with each observation. We wish to predict the expected loss for the next observable, based on the current data. If the predicted expected loss exceeds some value, we stop the process. Then

$$E_{Y|\mathbf{X}}(L(Y)|\mathbf{X}) = \int_0^\infty L(y)h(y|\mathbf{x})\ dy \tag{3.3}$$

A simple example of a loss function would arise in a warranty context. Say Y models the lifetime of the next component, which is warranted for a period of one time unit. If $Y < 1$, we have a loss of, say, 1, else we have no loss. The expected loss is, of course, the same as finding $P(Y < 1|\mathbf{X})$, where Y is the predicted next observation.

One can use any other loss function in a similar manner. We discuss two asymmetric loss functions which seem useful later in this chapter.

This Bayesian approach is particularly useful when one is indifferent to the model parameters, but very interested in controlling some loss. For example, when modeling skewed data, the practitioner might not care if the model used was a log-normal one or an inverse gaussian one. He might only care what his expected loss was. The advantage of the inverse gaussian model is that it allows the calculation of this expected loss much more easily than the lognormal model, using this predictive approach based on the sample.

3.1.2 Comparison with the self-starting scheme

Recall that we have used the same data sets as examples for both the self-starting and predictive charts. Comparing Tables 2.3 and 2.4 with Tables 3.1.1 and 3.2, we see that the predictive method almost detected the early outlier ($p = .9947$) and

was more sensitive than the self-starting technique (where $p = .9916$.) The predictive chart also signaled more strongly than the self-starting chart for the later outlier ($p = .9997$ for the predictive chart vs. $p = .9993$ for the self-starting scheme).

3.2 Sample of size m

A search of the literature does not reveal the predictive distribution for the next m observations, based on a sample of size n. In this section, we derive this distribution, using principles summarized by Geisser [1993]. We parallel the approach in Chhikara and Guttman, [1982], used for the univariate case.

The resulting distribution can be used in exactly the same manner as the preceding section for control of the process or the loss from a process.

For ease of computation, we use the parameterization of the density in the form $f(x|\theta, \lambda)$. We use the diffuse prior for θ and λ given by [Banerjee and Bhattacharyya, 1979], which is

$$p(\theta, \lambda) \propto \frac{1}{\lambda}$$

Then, the likelihood becomes

$$L(\theta, \lambda|\mathbf{x}) \propto \exp\left(-\frac{n\lambda v}{2} \left[1 + \frac{(\bar{x}\theta - 1)^2}{v\bar{x}}\right]\right)$$

where $\bar{x} = \sum x_i/n$ and

$$v = \frac{\sum(1/x_i)}{n} - \frac{1}{\bar{x}}$$

Note that we have slightly redefined v in this section, to match the notation of Chhikara and Folks.

The posterior density for the parameters becomes

$$f(\theta, \lambda|\mathbf{x}) = \frac{\left(\frac{\bar{x}}{\pi v}\right)^{1/2} \left(\frac{nv}{2}\right)^{n/2} \lambda^{(n/2-1)}}{\Gamma\left(\frac{n-1}{2}\right) S_{t,n-1}\left(\sqrt{\frac{n-1}{v\bar{x}}}\right)} \exp\left(-\frac{n\lambda v}{2}\left(1 + \frac{(\bar{x}\theta - 1)^2}{v\bar{x}}\right)\right) \tag{3.4}$$

We write the joint density of the future m observations, $\mathbf{y}$, given θ, λ as:

$$f(\mathbf{y}|\theta,\lambda) = \left(\frac{\lambda}{2\pi}\right)^{m/2} \prod_i^m y_i^{3/2} \exp\left(-\frac{\lambda}{2}\left(\sum_i^m \left(\frac{1}{y_i} - \frac{1}{\bar{y}}\right) + \frac{m(\theta\bar{y}-1)^2}{\bar{y}}\right)\right) \tag{3.5}$$

Then the predictive density becomes:

$$h(\mathbf{y}|\mathbf{x}) = \int_0^\infty \int_0^\infty f(\mathbf{y}|\theta,\lambda)\, p\,(\theta,\lambda|\mathbf{x})\, d\theta\, d\lambda \tag{3.6}$$

The simplification of $h(\mathbf{y}|\mathbf{x})$ is an extensive bit of work. We will use the following constants and redefined variables to reduce the notation. We will introduce additional notation later.

$$c = \frac{\left(\frac{\bar{x}}{\pi v}\right)^{1/2} \left(\frac{nv}{2}\right)^{n/2} (2\pi)^{-m/2} \left(\prod_i^m y_i^{-3/2}\right)}{\Gamma\left(\frac{n-1}{2}\right) S_{t,n-1}\left(\sqrt{\frac{n-1}{v\bar{x}}}\right)} \tag{3.7}$$

$$w = \sum_i^m \left(\frac{1}{y_i} - \frac{1}{\bar{y}}\right) \tag{3.8}$$

$$z = \frac{m\bar{x} + n\bar{y}(v\bar{x}+1) + w(\bar{x}\,\bar{y})}{\bar{x}\,\bar{y}} - \frac{(m+n)^2}{m\bar{y}+n\bar{x}} \tag{3.9}$$

Using the above change of variables and constants, we can write Equation 3.6 as:

$$h(\mathbf{y}|\mathbf{x}) = c \int_0^\infty \int_0^\infty \lambda^{\frac{m+n}{2}-1} \exp\left(-\lambda\left(\frac{z + m\bar{y} + n\bar{x}\left(\theta - \frac{m+n}{m\bar{y}+n\bar{x}}\right)^2}{2}\right)\right) d\theta\; d\lambda \tag{3.10}$$

Using the appropriate Gamma identity, we simplify to:

$$h(\mathbf{y}|\mathbf{x}) = c\Gamma\left(\frac{m+n}{2}\right) \int_0^\infty \left(\frac{z + m\bar{y} + n\bar{x}\left(\theta - \frac{m+n}{m\bar{y}+n\bar{x}}\right)^2}{2}\right)^{-(m+n)/2} d\theta \tag{3.11}$$

We make four new definitions:

$$\theta - \frac{m+n}{m\bar{y}+n\bar{x}} = \frac{t\sqrt{z}}{\sqrt{m+n+1}\sqrt{n\bar{x}+m\bar{y}}}$$

$$c_1 = c\,\Gamma\left(\frac{m+n}{2}\right) 2^{(m+n)/2}$$

$$c_2 = \frac{c_1\, z^{-(m+n-1)/2}}{\sqrt{m+n-1}\sqrt{n\bar{x}+m\bar{y}}}$$
$$LL = \frac{-(m+n)}{m\bar{y}+n\bar{x}}\sqrt{\frac{(m+n-1)(n\bar{x}+m\bar{y})}{z}}$$

Now, we change variables from θ to t. The integral for the predictive density becomes:

$$h(\mathbf{y}|\mathbf{x}) = c_1 \int_{LL}^{\infty} \left(z + \frac{zt^2}{m+n-1}\right)^{-(m+n)/2} \frac{\sqrt{z}}{\sqrt{m+n-1}\sqrt{n\bar{x}+m\bar{y}}}\, dt \tag{3.12}$$

Factoring out the z, grouping, and appealing to the symmetry of the integrand, we obtain:

$$h(\mathbf{y}|\mathbf{x}) = c_2 \int_{LL}^{\infty} \left(1 + \frac{t^2}{m+n-1}\right)^{-(m+n)/2} dt \tag{3.13}$$
$$= c_2 \int_{-\infty}^{LL} \left(1 + \frac{t^2}{m+n-1}\right)^{-(m+n)/2} dt \tag{3.14}$$
$$= c_2\, S_{t,m+n-1}(-LL) \tag{3.15}$$

Here, as before, the CDF of the Student's t-distribution with k degrees of freedom is given by $S_{t,k}$.

Now that the parameters have been successfully integrated out, the remaining task is to recover the variables of interest from the expression and simplify. We have:

$$\begin{aligned} h(\mathbf{y}|\mathbf{x}) &= c_2 S_{t,m+n-1}(-LL) \\ &= \frac{z^{-(m+n-1)/2}\Gamma((m+n)/2)2^{(m+n)/2}\prod y_i^{-3/2}}{\sqrt{(m+n-1)(n\bar{x}+m\bar{y})}\Gamma((n-1)/2)} \\ &\quad \times \frac{S_{t,m+n-1}(-LL)}{S_{t,n-1}\left(\sqrt{(n-1)/2}\right)} \left(\frac{\bar{x}}{\pi v}\right)^{1/2} \left(\frac{nv}{2}\right)^{n/2} (2\pi)^{m/2} \\ &= \sqrt{\left(\frac{\bar{x}}{(n\bar{x}+m\bar{y})v\prod(y_i^3)}\right)} \times \frac{S_{t,m+n-1}\left(\frac{(m+n)\sqrt{m+n-1}}{\sqrt{z(m\bar{y}+n\bar{x})}}\right)}{S_{t,n-1}\left(\sqrt{\frac{n-1}{2}}\right)} \end{aligned} \tag{3.16}$$

$$\times (nv)^{-n/2} \frac{\Gamma\left(\frac{n+m-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\pi^{m/2}} z^{-(m+n-1)/2} \tag{3.17}$$

$$= \sqrt{\frac{\bar{x}(nv)^n}{v}} \frac{\Gamma\left(\frac{n+m-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\pi^{m/2} S_{t,n-1}\left(\sqrt{\frac{n-1}{v\bar{x}}}\right)}$$

$$\times \frac{S_{t,m+n-1}\left(\frac{(m+n)\sqrt{m+n-1}}{\sqrt{z(m\bar{y}+n\bar{x})}}\right)}{z^{\frac{m+n-1}{2}} \prod_i^m y_i^{3/2}} \tag{3.18}$$

We write

$$z = \frac{n(v\bar{x}+1)}{\bar{x}} + w + \frac{m}{\bar{y}} - \frac{(m+n)^2}{m\bar{y}+n\bar{x}}$$

Notice that Equation 3.18 reduces to three summary statistics for the future observables: $w, \bar{y}$, and $\prod y_i$, which are functions of the harmonic, arithmetic, and geometric means, respectively. This contrasts with inference for the parameters, which reduced to the two sufficient statistics (which were the arithmetic and harmonic means). We can apply Equation 3.18 to our control charting scenario for groups of observations. Say we are sampling from a process which is in its start-up phase. We take samples of size m. After the first sample, we can compute a predictive density for the second sample. We find the probability of obtaining the current sample given the predictive distribution for the previous sample(s). If that probability is alarmingly low, we stop the process and investigate. Else, we incorporate the current sample into our "old" data and recompute a new predictive distribution, and sample again.

Determining the acceptance region of the future observables is best accomplished by use of the highest probability density approach, which results in a one-dimensional measure. We compute the value of the predicted density for the observations, and then compute $P\left(h(\mathbf{Y}|\mathbf{x}) > h(\mathbf{y}|\mathbf{x})|\mathbf{x}\right)$ using either Monte Carlo methods, or an indicator function and a numerical integration routine over m space.

In practice, this approach is more cumbersome than merely using the predictive format for a single future observable. This is especially true when attempting to

compute the predictive probability of the next set of observables. It does allow for fair treatment of all of the observations in the next sample, avoiding the temptation to order favorably or unfavorably the observations when applying the single observation prediction tests.

The multiple predictive density is of greater utility in computing expected losses for the next batch of the process, as opposed to the next sample. One can adjust the dimension of the future observables to any convenient size, and work with it. It might be that the appropriate size for prediction was, say, the output of the next shift at a plant, instead of predicting for merely the next sample or the quantities until the next sample.

3.3 Loss functions

There are two current loss functions widely in use in quality control.

The first is an indicator function for the region where the process is out-of-specification, $L(y) = 1_R(y)$. It is equal to one when the process is out-of-specification, and zero when the process is in-control. This is the implied loss function for many traditional charting schemes.

The second function in use is quadratic loss, whose best known proponent is Taguchi. Quadratic loss argues for continuous improvement to reduce variability, as any deviation from the target value for the process characteristic is penalized. Quadratic loss is symmetric with respect to deviations above and below target. Quadratic loss is also very tractable to work with, and is directly related to the variance of a process.

However, there is nothing sacred about quadratic loss as a choice of loss function, and upon reflection one can see that it does have modeling weaknesses.

It is a large assumption that one's losses are quadratic in form. Most are

asymmetric. For example, consider filling a bag with flour. The loss for slightly over-filling the bag is likely much less than the loss for under-filling, with its associated losses of good-will and regulatory penalties. Similarly, consider the tensile strength of 550 pound parachute cord. The loss for making the cord too strong is probably much smaller than the loss for making the cord too weak, and the loss for making the cord too weak grows (in this parachutist's opinion) much faster than quadratic growth.

Loss functions ideally would come from process understanding. Many of the losses for being out-of-control are intangible, or poorly estimable. That lack of understanding of the loss has argued for a quadratic loss function, which at least is easy to work with. However, it is also possible to work with at least two other types of asymmetric loss functions.

3.3.1 Lin-quad loss

This loss function is a piecewise differentiable function which is linear on one side of the target value and quadratic on the other. It is easy to derive its form. Let the quadratic loss be specified by $L_1(y) = k(x-c)^2$, where c is the target value and k is the coefficient. Let the linear loss be specified by $L_2(y) = b(x-d)$. d adjusts the intercepts, as we will see below. To obtain differentiablity, we define the point $x = b/(2k) + c$ as the point where the definition changes. We have then:

$$L(y) = \begin{cases} b(x - \frac{b}{4k} - c) & \text{for } x < c + \frac{b}{2k} \\ k(x-c)^2 & \text{for } x \geq c + \frac{b}{2k} \end{cases} \tag{3.19}$$

Here we have placed the linear portion on the left side ($b < 0$). A similar procedure works for the linear loss on the right hand side. This loss function is pictured in Figure 3.1.

For the multivariate predictive case, one way we can obtain our multivariate loss is by adding the marginal expected losses.

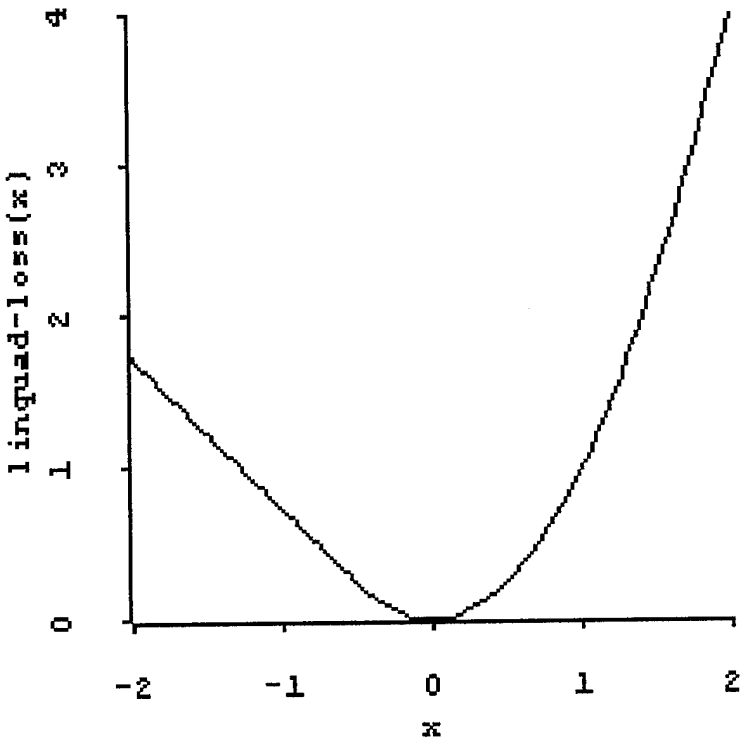


Figure 3.1: Linear-quadratic loss function for target value 0.

The coefficients c, k and b can be estimated in the usual way from historical loss data or assumed as part of the modeling effort.

3.3.2 LINEX Loss

Varian [1975] proposed, Zellner [1986] further investigated, and Geisser [1993] used an asymmetric loss function called LINEX. Varian defined $\Delta = \hat{\theta} - \theta$ as the scalar estimation error. His loss function was then

$$L(\Delta) = b \exp(a\Delta) - c\Delta - b$$

This loss function is illustrated in Figure 3.2.

This loss function is recommended where the consequences for error to one side of the target value are catastrophic. Geisser [1993] discusses this loss function in the context of regulating insulin dosage to diabetics, where too much insulin is catastrophic, compared to too little.

We can use a loss function of similar form for our predictive control problem, where $\Delta = y - c$, and c is our target value. Again, we can find the expected loss by integrating against the predictive density.

3.4 Conclusions

The advantages of using a predictive method should now be clear: first, we obtain tighter predictions; second, we can integrate our predictive density against a loss function and obtain our expected loss. We can base our decisions on this value. By further incorporating in the cost of stopping the process to investigate a signal, we can act on our total expected loss. This offers a choice: instead of acting on possible parameter shifts which may cost more to investigate than to tolerate, we can decide on economic grounds.

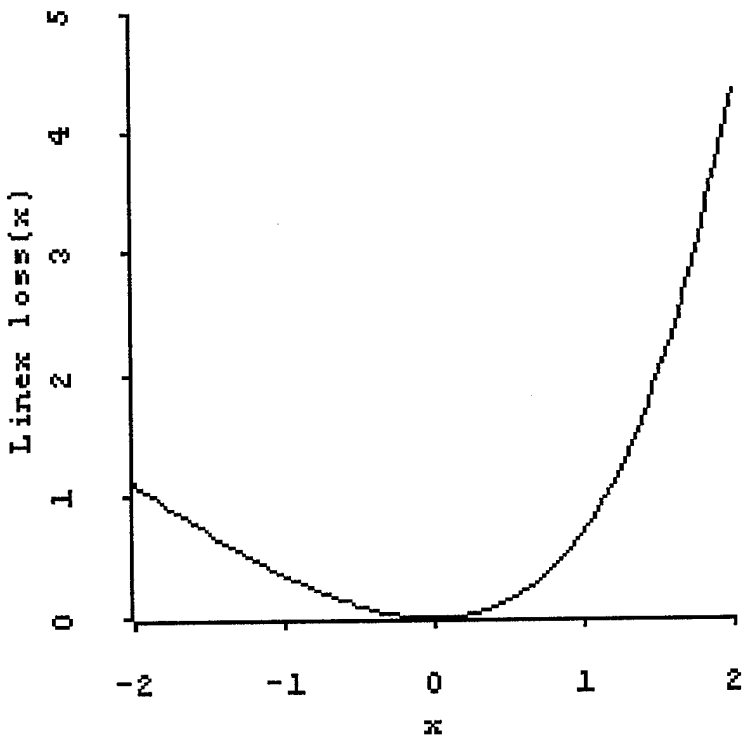


Figure 3.2: Linear-exponential loss function for target value 0.

Chapter 4

Cumulative Sum Charts for IG Processes

In this chapter, we derive a CUSUM scheme for processes modeled by the IG distribution. We chart for location and shape, and then examine the behavior of the tests under the model, and under various model departures.

4.1 Scheme construction

Control charts for normal processes chart location and scale. The statistic for location is usually the sample mean; the one for scale is usually the sample standard deviation. For the IG distribution, those statistics (sample mean, sample standard deviation) are not independent. To use them to control an IG process, one would need a bivariate chart (discussed below). However, the control limits for such a chart would seem artificial, as indicated by the plot of $\bar{X}$ vs. S for an IG example in Fig 4.1, below. A better scheme for joint control will be discussed later.

We return to well-known principles to derive our CUSUM scheme. Recall from Chapter 1 that the inverse gaussian distribution is a member of the exponential family. As discussed in Hawkins [1992c] and in Chapter 1, the upward CUSUM for a member of exponential family in decision interval form is defined by

$$S_0^+ = 0 \tag{4.1}$$

$$S_n^+ = \max(0, S_{n-1}^+ + T_n - k^+) \tag{4.2}$$

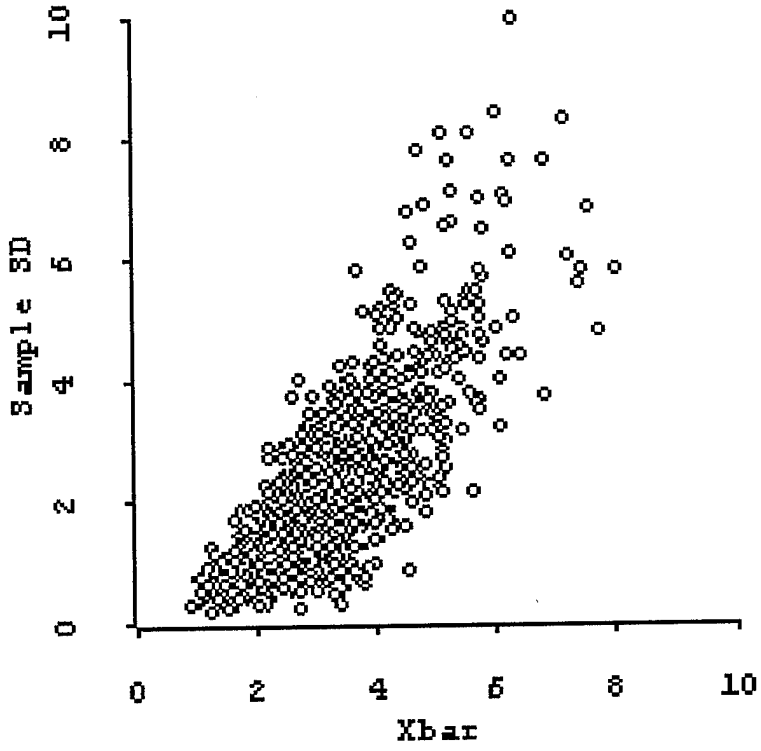


Figure 4.1: A plot of $(\bar{X}, S)$ for samples of size 5 from the $IG(3,5)$. Notice that as $\bar{X}$ increases, so does the magnitude and dispersion of S.

$$k^+ = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)}, \tag{4.3}$$

where $b(\theta)$ and $d(\theta)$ come from the parameterization of the exponential family as $f(x) = exp(a(x)b(\theta) + c(x) + d(\theta))$ and $T_n = a(X_n)$.

Similarly, the downward CUSUM is given by:

$$S_0^- = 0 \tag{4.4}$$

$$S_n^- = \min(0, S_{n-1}^- + T_n + k^-) \tag{4.5}$$

$$k^- = \frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)}$$

We assume as our model that our process is well specified as $X \sim IG(\mu, \lambda)$. For two parameters, we write the density of X as

$$f(x; \phi_1, \phi_2) = \sqrt{\frac{\phi_2}{2\pi}} \exp\left(\sqrt{\phi_1\phi_2}\right) x^{-3/2} \exp\left(\frac{-1}{2}(x, 1/x) \cdot (\phi_1, \phi_2)\right)$$

where $\phi_1 = \frac{\lambda}{\mu^2}$ and $\phi_2 = \lambda$.

We will use four charts simultaneously: one each for upward and downward departure in the two model parameters.

The schemes derived below are for individual observations, or samples of size one. This has the advantage of maximum flexibility: one can always chart a larger sample as a group of individual observations. While we do not do so here, it is possible also to derive schemes for samples larger than size one by considering CUSUMs of the $\bar{X}$ and V, the minimal sufficient statistics for the sample.

4.2 CUSUMs for location

We use the decision interval form for the CUSUM, as it is more easily implemented by computer. We maintain a pair of CUSUMs for location; one for upward departures

from the mean and the second for downward departures, denoted by S^+ and S^-, respectively.

For our CUSUM for location, we assume that $\phi_2 = \lambda$ is fixed and known. We then have a one-parameter exponential family in ϕ_1. To determine our settings for the CUSUM chart, we first decide for which shifts of the mean we wish maximum sensitivity. We call the upper value μ^+ and the lower value μ^-. The corresponding values of ϕ_1 are called ϕ_1^+ and ϕ_1^-.

With the shape parameter fixed, we have

$$f(x, \phi_1) = \exp(a(x)b(\phi_1) + c(x) + d(\phi_1))$$

with $a(x) = x$, $d(\phi_1) = \sqrt{\phi_1 \lambda}$ and $b(\phi_1) = -\phi_1/2$. Then the reference value, k_i, is given by Equation 4.3, which simplifies to

$$k_i = -\frac{-2\sqrt{\lambda}}{\sqrt{\phi_i} + \sqrt{\phi}} \tag{4.6}$$

where $\phi = \lambda/\mu^2$ and ϕ_i is ϕ_1^+ or ϕ_1^-, as appropriate.

Rewriting Equation 4.6 in terms of the parameters μ and λ, we obtain

$$k\ (\mu_i) = \frac{2\mu\mu_i}{\mu + \mu_i} \tag{4.7}$$

This value for k is the *harmonic mean* of the in-control and out-of-control parameters. Notice that this value for k is very close to the heuristic value one would obtain by using the arithmetic mean. The difference between the heuristic value and the optimal value for k is only

$$\Delta = -\frac{(\mu_0 - \mu_1)^2}{2(\mu_0 + \mu_1)} \tag{4.8}$$

The CUSUM scheme for location then consists of two decision interval charts, one for upward detection and one for downward detection. They are given by Equations 4.1, 4.2, and 4.7 for the upward CUSUM and Equations 4.4, 4.2, and 4.7 for

the downward CUSUM. In both cases, $T_n(X_n) = X_n$. This scheme has the advantage of simplicity: it accumulates the sum of the observations themselves less k, without the need for any transformation. The k term depends only on the mean under the null and alternative hypothesis. The parameter ϕ_1 is not as simple as the mean or variance of the process, but it does have the characterization as the process mean divided by the process variance:

$$\phi_1 = \frac{\lambda}{\mu^2} = \frac{\mu}{\frac{\mu^3}{\lambda}} = \frac{EX}{Var(X)}$$

also known as the process signal-to-noise ratio.

The scheme signals when either $S_n^+ > h^+$ or $S_n^- < h^-$.

h^+ and h^- are set by considering the average run length, that is, the average number of samples until $S_n^+ > h^+$ or $S_n^- < h^-$, resulting in a false signal when the process remains in control. These average run lengths are discussed below.

4.3 CUSUMs for shape

To derive our CUSUM for the shape parameter, we assume μ fixed and known, and again consider the resulting one parameter exponential family. We take

$$a(x) = \frac{(x-\mu)^2}{x\mu^2}$$

$b(\lambda) = -\lambda/2$, and $d(\lambda) = \ln(\lambda)/2$. The sequential probability ratio test for a shift from λ to λ_0 involves summing $a(x_i)$ until it exceeds a limit which depends on the number of terms in the sum.

We could attempt to follow the same procedure as in the previous section to construct our CUSUM for shape. This would result in the scheme $S_0^+ = 0$, $S_n^+ = \max(0, S_{n-1}^+ + a(X_n) - k)$, with

$$k = \frac{d(\lambda_0) - d(\lambda)}{b(\lambda_0) - b(\lambda)} = \frac{\ln(\lambda_0/\lambda)}{\lambda_0 - \lambda} \tag{4.9}$$

We prefer to focus instead on the fact that $\lambda a(X) \sim \chi^2_1$ and to derive a CUSUM scheme for that distribution. Then a test for a shift in λ becomes a test of a scale change in a χ^2_1, or, better, a scale change in a $\Gamma(1/2, 2)$ distribution, with the gamma density given in the form

$$f(x; \alpha, \beta) = \frac{\beta^{-\alpha} x^{\alpha-1} \exp(-x/\beta)}{\Gamma(\alpha)} \tag{4.10}$$

By focusing on the more general problem of a scale shift for a Gamma distribution, we get two CUSUM schemes for the price of one.

Note that the Ph.D. dissertation of Regula [Regula, 1976] considered a shift in the shape parameter, α, for a Gamma distribution. A search of the literature does not reveal an explicit CUSUM test for the scale parameter, β, of a Gamma distribution, although it follows easily from the CUSUM for the variance of a normal distribution derived by Johnson and Leone [1961b], and the work of Hawkins [1992c].

Let $Y \sim \Gamma(\alpha, \beta)$. The log-likelihood ratio for the Gamma distribution for a test of $\beta = \beta_a$ against $\beta = \beta_0$ using Equation 4.10 simplifies to

$$\Lambda_n = \sum_1^n -\alpha \ln \frac{\beta_a}{\beta_0} + \frac{(\beta_0 - \beta_a) y_i}{\beta_0 \beta_a} \tag{4.11}$$

Following our pattern of using the SPRT to obtain CUSUM schemes, Equation 4.11 motivates the following CUSUM scheme for the shape parameter of the gamma distribution:

$$\begin{aligned} S_0 &= 0 \\ S_n &= \max(0, S_{n-1} + Y_n - k) \\ k &= \frac{\alpha \beta_0 \beta_a \ln(\beta_0/\beta_a)}{\beta_a - \beta_0} \end{aligned} \tag{4.12}$$

As usual, the scheme signals when $S_n^+ > h$, where h is found by determining the *a priori* desired ARL.

A FORTRAN routine for determining h for the scale change for a $\Gamma(\alpha, \beta)$ is available from the author.

For the problem of the shape parameter given $X \sim IG(\mu, \lambda)$, we note that the distribution of

$$\frac{\lambda(X-\mu)^2}{X\mu^2} = \lambda a(X) \sim \chi_1^2$$

We see that a shift in λ results in a scale shift in a $\Gamma(1/2, 2)$ and can be detected using the scheme just derived for the scale shift of a Gamma distribution.

Our in-control parameter for β is 2. When the true value of λ shifts to λ_a, the distribution of $\lambda_0 a(X)$ shifts to

$$\Gamma\left(\frac{1}{2}, \frac{2\lambda_0}{\lambda_a}\right)$$

So our alternate point hypothesis for the CUSUM of the Gamma scale parameter becomes $\beta = \frac{2\lambda_0}{\lambda_a}$.

Our CUSUM scheme then becomes

$$\begin{aligned} S_0^+ &= 0 \\ S_n^+ &= \max(S_{n-1}^+ + \lambda_0 a(X_n) - k) \qquad (4.13) \\ k &= \lambda_0 \frac{\ln \frac{\lambda_0}{\lambda_a}}{\lambda_0 - \lambda_a} \qquad (4.14) \end{aligned}$$

This scheme is illustrated in Table 4.1 and its behavior is explored in detail in the next section.

We note that it is the same scheme as the one we derived from first principles in Equation 4.9 but has more general application to the problem of scale shift for an arbitrary Gamma distribution.

μ_0	μ_1	λ	h	ARL in control	ARL out-of-control
3	3.5	5	1	4.742	3.639
			5	16.340	9.730
			10	44.877	20.314
			20	178.354	47.989
			40	1233.208	115.569
μ	λ_0	λ_1	h	ARL in control	ARL out-of-control
3	5	4	5	31.691	17.446
			10	105.302	40.639
			20	532.991	101.779
			40	5163.017	239.129

Table 4.1: Some in-control and out-of control ARL values for various CUSUM parameters. Out-of-control values are taken for the parameter at the alternate (tuning) value.

4.4 ARLs in control

We provide two methods for computing the ARLs of these schemes. The first, due to Jun and Choi [1993] uses simulation and variance reduction techniques. The second uses an approximation due to Hawkins [1992c] which approximates the underlying integral equations to find the ARL. Table 4.1 provides some typical in-control and out-of-control ARL values. Details of algorithms and coding are in the Appendix.

The variance reduction scheme was used as an independent check on the author's programming implementation of the faster, more accurate integral approximation.

Below we provide graphs of h versus $\ln(ARL)$ for the CUSUM chart in-control and out-of-control at the alternate value. Similar charts could be used to select h. Additionally, we provide FORTRAN routines in the appendix for finding h for any ARL for both μ and λ, given the parameters of the process in- and out-of-control.

It is interesting to note from the graphs that the rate of increase in ARL versus h appears to be approximately exponential for the in-control case and linear for the out-of-control case.

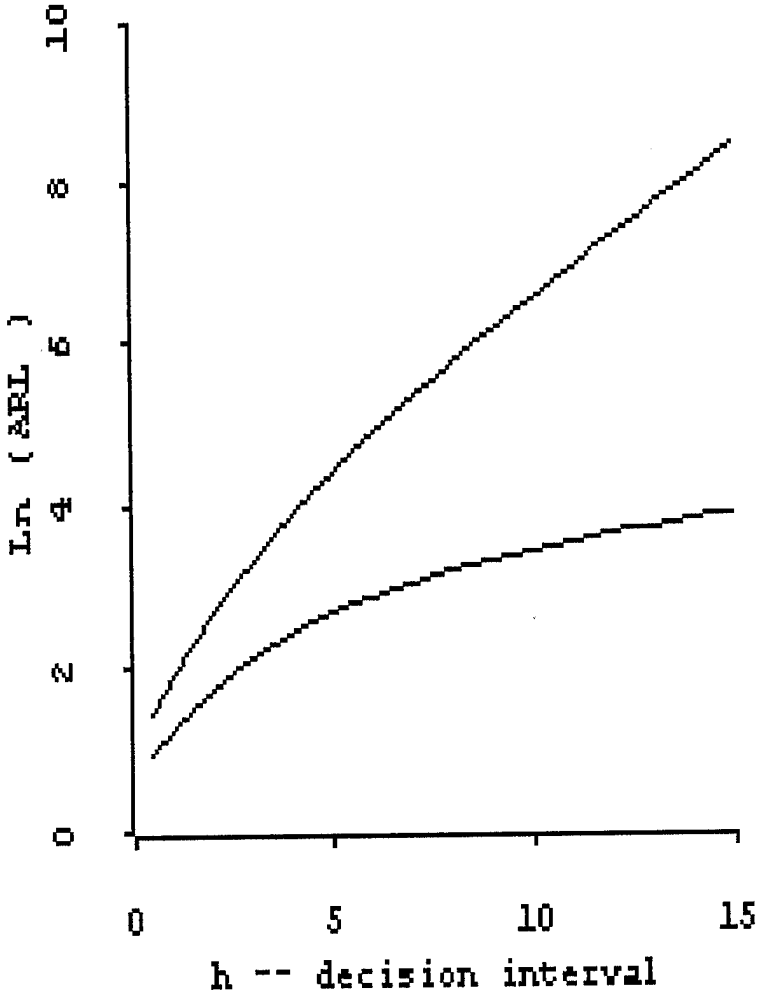


Figure 4.2: A chart of h vs. $\ln ARL$ for the CUSUM for detecting shifts in μ, with $\mu_0 = 3$, $\mu_1 = 3.5$, and $\lambda = 5$. Here we consider samples of size 5. The lower curve is the $\ln(ARL)$ for the out-of-control state.

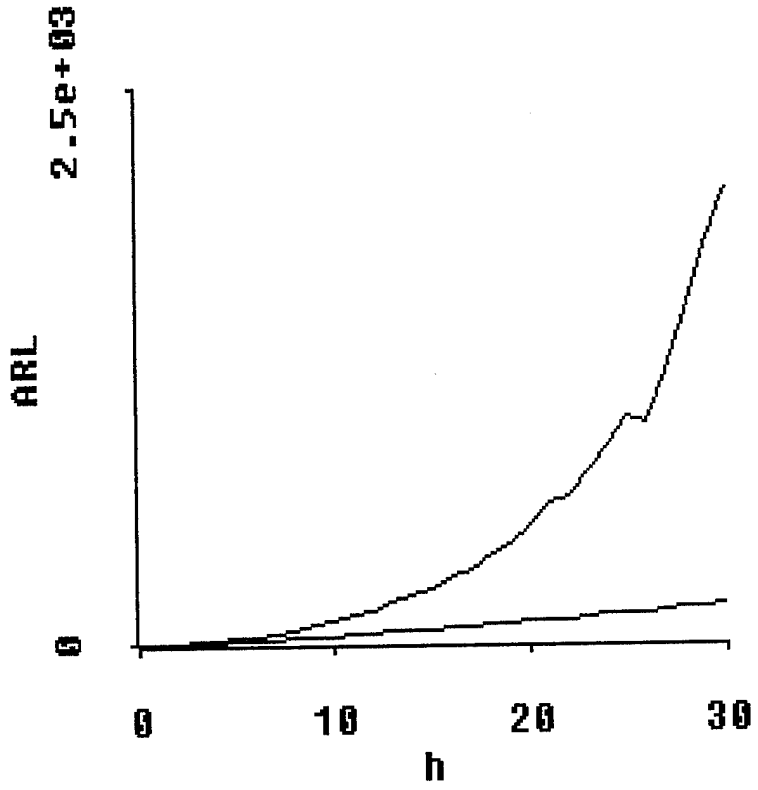


Figure 4.3: A chart of h vs. ARL for the CUSUM for detecting shifts in λ, with $\mu = 3$, $\lambda_0 = 5$, and $\lambda_1 = 4$, with samples of size 5. The lower curve is the ARL for the out-of-control state. There is an anomaly for $h = 26$, from the FORTRAN output, causing the inappropriate dip in the curve. Note that this figure in not in logarithmic scale.

4.4.1 Two pedagogical notes

We mention in passing that the algorithm for finding h to meet a specified ARL is a straightforward application of the Newton-Raphson root finding algorithm, with one small twist: the derivatives are approximated numerically, instead of being computed symbolically. The routine is very fast, usually producing convergence in a handful of steps.

We also mention that the variance reduction scheme for the simulation method is accessible to the introductory (calculus-based) statistics student and provides a nice motivation for discussing covariance.

4.4.2 Comments on ARL as a measure of effectiveness of a CUSUM scheme

There have been objections in the literature to using ARL as the sole measure of the effectiveness of a scheme [Gan, 1992] [Bissell, 1969][Barnard, 1959]. If the distribution of the run-length for the CUSUM was known for a particular scheme, it would clearly be preferred for characterizing a CUSUM scheme. The run-length distribution is known for some simple distributions, such as the exponential [Gan, 1992a]. However, in the absence of knowledge of the run-length distribution (such as in the case of the inverse gaussian distribution), ARL remains the accepted method for evaluating the performance of CUSUM schemes. The practitioner is advised to keep in mind that the run-length is often skewed, and that the median run-length may differ sharply from the average run-length. Deriving the exact run-length distribution for the inverse gaussian CUSUM scheme is a topic for later research. Approximating the run-length distribution can be done easily by simulation.

We recommend that the practitioner present simulation results indicating the general shape of the run length distribution for a scheme, as well as the ARL, when

briefing a designed CUSUM scheme to the client.

4.5 Performance for small persistent shifts

We consider the performance of the CUSUM schemes for detecting small, persistent shifts in the parameters. Moustakides [1986] proved that the schemes are optimal in the sense they have the smallest expected number of samples until a signal, when shifting out of control to the alternate value, of all schemes with similar in-control false alarm rates. We construct tables to illustrate the response rates.

4.5.1 Small persistent shifts in μ

In the following tables, we present the performance of the CUSUM scheme in detecting small persistent shifts. We assume that the process is exactly specified a priori.

We desire some benchmark to examine just how optimal is the optimal procedure. For comparison, we use the CUSUM scheme for the mean of a normal distribution, which uses the arithmetic mean between the parameter values in-control and out-of-control as its reference value. Recall that our optimal CUSUM for the mean of an Inverse Gaussian random variable uses the harmonic mean between the in-control and out-of-control values. Since the harmonic mean is close to (but strictly less than) the arithmetic mean for small shifts, this should provide a reasonable competitor. We will call the use of the arithmetic mean for k for the IG case a "naive" CUSUM scheme, and the use of the optimal harmonic mean the "optimal" scheme.

We tabulate the out-of-control ARLs for both schemes for some illustrative values of μ, λ, and ARL in-control. We use the one-sided CUSUM in these tables. Similar results hold for the two-sided CUSUM.

Out-of-control ARL					
Parameters					
μ	μ_a	λ	ARL in-control	Naive CUSUM	Optimal CUSUM
3	3.5	5	100	34.250	34.247
3	2.5	5	100	34.313	34.299
3	5	5	100	11.429	10.072
3	3.5	10	100	25.690	25.683
3	2.5	10	100	24.581	24.563
3	3.5	5	1000	106.981	106.894
3	2.5	5	1000	94.902	94.730
3	3.5	10	1000	68.469	68.388
3	2.5	10	1000	24.581	24.563
10	11	10	100	54.806	54.806
10	9	10	100	58.301	58.300
10	20	10	100	9.808	9.734
10	11	100	1000	243.701	243.688
10	9	100	1000	241.259	241.229

Table 4.2: Comparing performance by ARL of various control schemes to detect a small persistent shift in the mean.

Shewhart one-sided ARLs were calculated as

$$ARL = \frac{1}{P(X > crit | X \sim IG(\mu_a, \lambda))} \tag{4.15}$$

by numeric integration.

The naive and optimal CUSUM schemes were designed for their in-control ARLs. h was found using the FORTRAN routines in the Appendix. k depends, of course, on the scheme.

We see from Table 4.2 that for small shifts the use of the optimal harmonic mean as a reference value does beat the benchmark arithmetic mean for performance, but not by much. This is not surprising, given that Equation 4.8 indicates that values of the harmonic mean and arithmetic mean are very close for small shifts.

4.5.2 Small persistent shifts in λ

We will repeat the process of the previous section to obtain a benchmark for the CUSUM for λ. We construct a naive alternative for k by averaging the expected values of $a(X)$ when the process is in-control then out-of-control. We recognize that

Out-of-control ARL			
Parameters			
λ	λ_a	ARL in-control	Optimal CUSUM
10	11	100	70.03
10	9	100	61.08
10	11	1000	442.12
10	9	1000	373.09
5	5.1	100	92.28
5	4.9	100	90.14
5	5.1	1000	777.17
5	4.9	1000	746.67
100	101	100	96.01
100	99	100	94.91
100	101	1000	875.24
100	99	1000	860.24

Table 4.3: Out-of-control ARLs for a small persistent shift in λ. Note that a small persistent change in λ can be very difficult to detect.

the expected value of $a(X)$ in control is 1, and that when $\lambda = \lambda_a$, the expected value of $a(X)$ is $E(a(X)) = \frac{\lambda}{\lambda_a}$. Averaging those two expected values results in $k = 1/2 + \frac{\lambda}{2\lambda_a}$.

These two values for k are very close. For example, when $\lambda = 10$ and $\lambda_a = 11$, the two values for k differ by only .00144365.

We construct Table 4.3 to explore small persistent shifts in the scale parameter, λ. We see that it is very difficult to detect a small persistent shift in λ.

Examination of Table 4.3 shows that the optimal scheme has shorter out-of-control ARLs than the benchmark. It also shows that the optimal scheme in not greatly more powerful than the benchmark. This is a result of the underlying robustness of the CUSUM to mis-specification of the out-of-control state when designing the CUSUM scheme. One could look at the benchmark value here as the optimal value to detect some different shift in the mean, and see that the performance was still fairly close to the scheme designed for the actual shift in the mean.

4.6 Conclusions

In this chapter, we derived the optimal tests for the CUSUM of the inverse gaussian distribution for both the mean and shape parameter. In the process of doing this, we developed an optimal test for shift of scale for the CUSUM of a χ^2_ν random variable. We found an immediate application for that χ^2_1 test in our test for the shape parameter. We adapted existing FORTRAN codes to evaluate the ARL of these tests. We checked those FORTRAN codes against a variance reduction simulation scheme. We compared the optimal tests against the usual heuristic and found that for small changes in the mean the optimal test was not significantly better than the usual heuristic.

We developed codes which allow the interested party to design an optimal or traditional CUSUM scheme, and to find its average run length.

Chapter 5

CUSUM Embellishments for IG processes

In this short chapter, we address two CUSUM embellishments and how they apply to the work in the previous chapter.

5.1 Fast initial response CUSUM

Lucas and Crosier [1982] proposed the fast initial response (FIR) CUSUM. They reasoned that if the CUSUM was started not at zero, but at some value between 0 and the decision-interval value h, the CUSUM would respond faster to early out-of-control states. However, if the process was in-control in those early states, the CUSUM would most likely return to zero, and nothing would be lost.

Lucas and Crosier recommend a head start value for $S_0 = h/2$, based on numerical evidence for the normal scheme.

The FORTRAN programs discussed in the previous chapter can take advantage of a feature of the CUSARL code of Hawkins. That code reports the FIR ARL for $S_0 = h/2$ along with the regular ARL. It is, therefore, simple to adjust the programs which find the ARL for the $IG(\mu, \lambda)$ to also find the FIR ARL. Generally, one pays for increased speed of detection when starting out-of-control with a slightly lower ARL when starting in-control. To maintain the same in-control ARL, one adjusts h upward appropriately, until one obtains the desired in-control FIR ARL. Of course, if the process is initially in-control, but goes out-of-control later, there will be a delayed

response compared with the regular ARL as the CUSUM moves toward the higher value of h.

One should use the FIR, then, if one is concerned that the process goes out-of-control early. One obtains that increased detection power at the expense of decreased power to detect later departures from control.

While the ARL may not be significantly affected by the use of the FIR scheme, the underlying run-length distribution is. One incurs more frequent short-length runs, accentuating the already skewed nature of the run-length distribution. This could be of no small annoyance to the process manager, who has already had to learn that the mean run length is greater than the median run length.

One context where such a trade-off is desirable is immediately after repairing the process due to an earlier signal. If one has not correctly diagnosed and repaired the process, the process is still out of control. Upon restarting the process, one would wish to know this quickly, to avoid continuing out of control.

The practice of using FIR CUSUMs is not universally accepted, because of the above trade-offs.

A topic for additional research would be to determine an algorithm for finding the optimal head start value to meet some criteria. Since one obtains the ARLs for many states when one uses the Markov-chain approximation for finding ARLs (discussed in the appendix when the CUSARL code is presented), one could either determine a better rule to select a head start or validate the $h/2$ heuristic.

5.1.1 An example

Let's return to Figure 4.2, which illustrated $\ln(ARL)$ versus h for an example. In Figure 5.1, we add the FIR scheme for the same parameter values. This means we start at $S_0^+ = h^+/2$ and $S_0^- = h^-/2$. We see that the out-of-control ARL is greatly

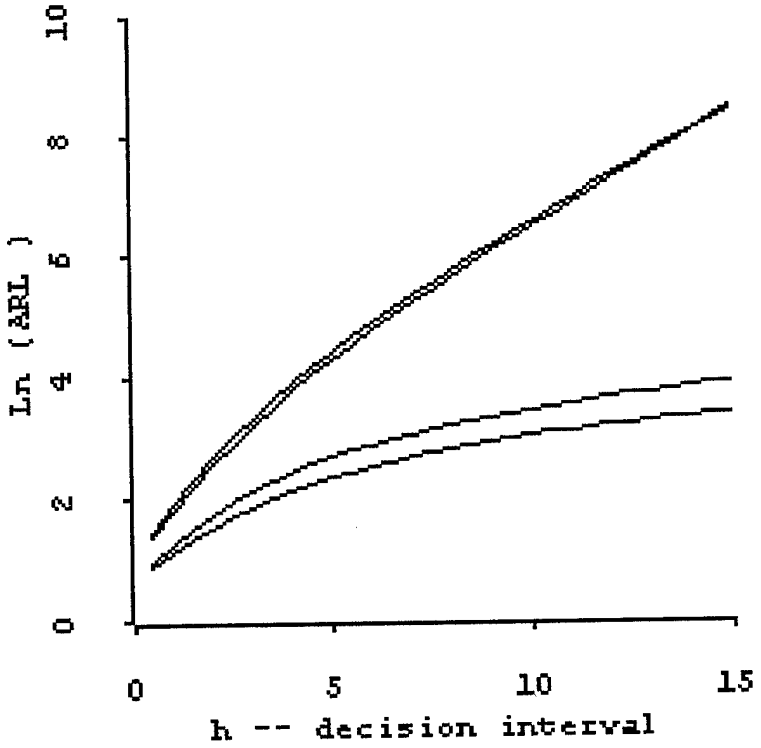


Figure 5.1: Comparison of FIR and regular ARL schemes, with $\mu_0 = 3.0$, $\mu_1 = 3.5$, $\lambda = 5$, and $n = 5$, where n is the size of the rational subgroup. The FIR response is the lower of each pair of curves; the upper pair the in-control ARL and the lower pair the out-of-control ARL.

Table 5.1: Table of comparable values for regular and FIR CUSUMs, with $\mu_0 = 3.0$, $\mu_1 = 3.5$, $\lambda = 5$, and $n = 1$, where n is the size of the rational subgroup.

Value	Regular CUSUM	FIR CUSUM
k	3.2307	3.2307
h	37.5619	38.8170
S_0	0	19.4085
ARL in-control	1000	1000
ARL out-of-control	106.89	75.2727

reduced, while the in-control ARL is not affected nearly so much. However, if the CUSUM returns to zero (a regeneration point), the ARL then becomes much greater that the standard CUSUM scheme because of the higher value of h, the decision value.

As a more focused example, we look at the design parameters for an ARL of exactly 500 for this scheme. Using a routine available from the author, we obtain the information in Table 5.1.

If the FIR CUSUM returns to zero before reaching h, the ARL for the scheme in Table 5.1 becomes 1114.7690 in-control. The out-of-control ARL increases as well, to 111.3589.

Accordingly, we see the fast response feature of the FIR only applies if the CUSUM goes out of control before returning to zero. If the CUSUM does return to zero, the scheme performs worse than the regular ARL. One should then use the FIR with caution. The FIR is still a good choice for starting the scheme after a signal if there is doubt as to whether or not the condition causing the signal has been corrected.

5.2 Self-starting CUSUM for the mean

The arguments in Section 2.4.2 for self-starting charts for the Shewhart Control scheme apply with greater strength to the CUSUM scheme, which was their original context. In fact, our attention was first drawn to self-starting schemes in the article by Hawkins [1987], who specifically addressed self-starting CUSUM charts of normal variates.

We will use the same transformations in the self-starting CUSUM scheme that we used in the self-starting Shewhart scheme for the mean. That is, we will CUSUM

$$Z = \Phi^{-1}(F_T(T)) \sim N(0,1)$$

with $\Phi^{-1}(z)$ being the inverse CDF for the standard normal variate. Recall

$$T = \frac{\sqrt{n_1 n_2 (n_1 + n_2 - 2)}\ (\bar{X} - \bar{Y})}{\sqrt{\bar{X}\,\bar{Y}(n_1\bar{X} + n_2\bar{Y})(V_1 + V_2)}}$$

.

Our self-starting CUSUM scheme, in control, becomes a CUSUM of $N(0,1)$ variates.

The sole difficulty becomes the computation of the reference value, k. The distribution of T_i when there is a model departure is not known, depends on the length of time the process has been running in control, and appears likely to be too complicated to be worth the effort to find it.

We note that Hawkins [1987] finessed the issue in his paper. He expressed the shift of the mean in the original variable as a multiple of standard deviations, and used one–half that value as the reference value. Since his intermediate studentized residuals can be thought of as approximating the number of standard deviations away from the unknown mean, this seems reasonable. The additional transformation to the $N(0,1)$ distribution does not radically affect this heuristic. This approach avoided the difficulty of finding the k in the context of the transformed variables.

Since we, by contrast, are using an exact transformation to normal variables, we fall back on standard procedure. We will set k as being the average between the expected value of Z in control and the expected value of Z when the process is out of control at some level. When the process is in control, the expected value of Z is 0 and the standard deviation of Z is 1. We appeal to the robustness of the CUSUM reference value and select the out-of-control mean value to be 0.1. This value is obtained from a simulation of various out-of-control states at various points, computing the value of Z for the first observation out-of-control. While it is a very rough approximation, the CUSUM is known to be robust to mis-specifications of the out-of-control state. This results in a reference value of $k = 0.05$. When the underlying IG process goes out of control, both the mean and the standard deviation of the Z change. Depending on how long the process has been running, these will change either slowly or rapidly back to $N(0,1)$ as the process fails to signal over time. However, if the process has been running an appreciable time, the drift back to the original expected value should be relatively slow.

If we don't know the distribution of the summand for the CUSUM when the process is out of control, it is not possible to analytically derive an optimal CUSUM scheme. If we were concerned with finding a better reference value for a given out-of-control state, we could simulate to approximate the expected value of $\Phi^{-1}(F_0(T))$, and then use the rule for shifts of means of normal variates which sets

$$k = \frac{\mu_o + \mu_1}{2}$$

Since under a model shift, $F_0(T)$ is no longer distributed uniformly and it follows that $\Phi^{-1}(F_0(T))$ is not normal, this is at best an approximation.

However, we are proposing this self-starting CUSUM as a reasonable, not optimal, scheme. We defer further discussion of optimality for subsequent work, and proceed to examples.

As an example, consider the following scenario. Let our true (but unknown) process parameters be $\mu = 42$, $\lambda = 66$. We will look at samples of size 1, and at observation 50 we will change the distribution to $IG(52, 66)$, corresponding to a moderate shift in the process.

The self-starting CUSUM is illustrated in Figure 5.2. We run in-control for 50 cases with an $IG(42, 66)$. We will go out of control at observation 51 and beyond, moving to an $IG(52, 66)$. We set our $ARL = 100$. Since we have set $k = .05$, we have $h^+ = h^- = 10.2969$. In this case, the self-starting CUSUM signaled at the 66th observation, or 16 observations after the process went out of control.

For short start-ups before going out of control, it is not unusual for the scheme to fail to detect the change. The example above can be considered a short training set, since 50 points corresponds to only 10 samples of size 5.

Performance is better for longer training sets, as indicated in Figure 5.3.

Note from Figure 5.3 that, even with the long start-up of 150 observations, the self-starting scheme begins to adjust to being out-of-control. This can be seen from the CUSUM for S^+, which starts to tail back to the horizontal center line after observation 225 or so, despite remaining out-of-control.

For the behavior of the out-of-control CUSUMs, including run-length distribution, we must resort to simulation, since the CDF for the distribution of the out-of-control Z is not known.

We reiterate that the ARL will depend not only on the size of the shift, but the length of time the process has been running in-control prior to going out of control.

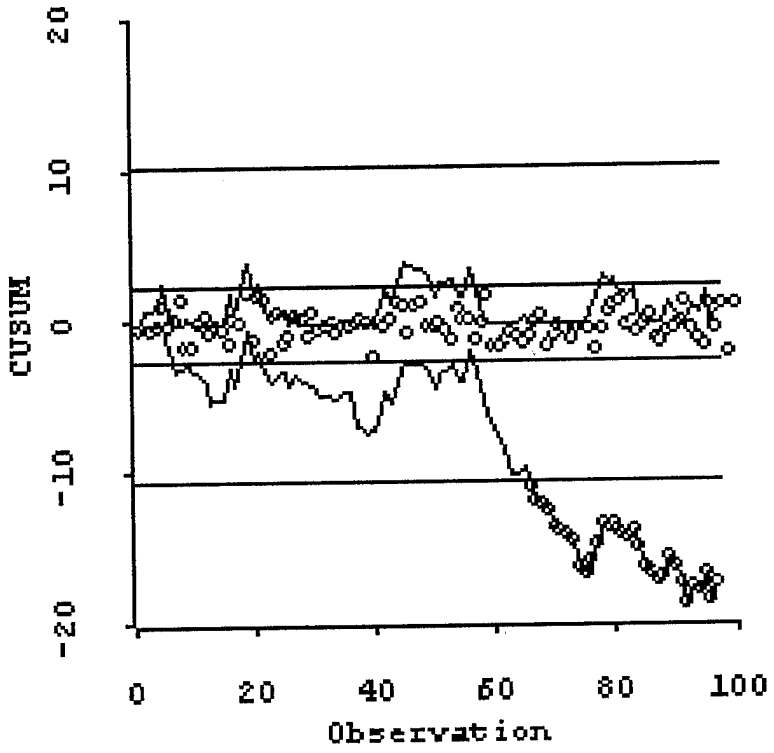


Figure 5.2: A combined self-starting Shewhart and self-starting CUSUM chart. The process ran in-control ($IG(42, 66)$) for 50 samples of size 1, then went out of control to $IG(52, 66)$. The chart signals at observation 66 for the CUSUM, and observation 80 for the Shewhart.

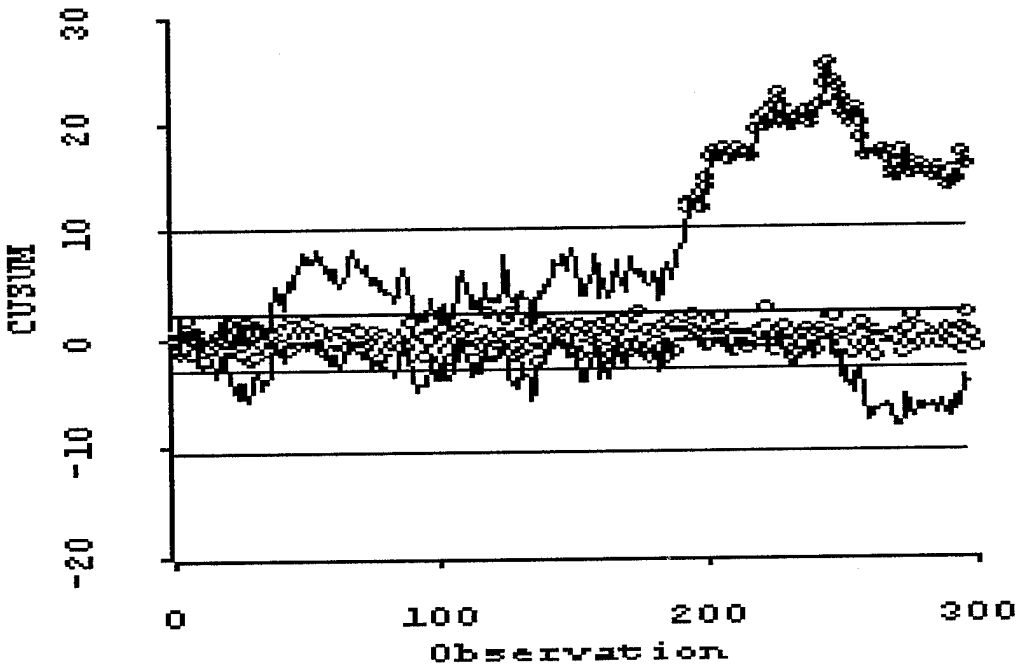


Figure 5.3: Self-starting CUSUM for the mean of an IG random variable. The process is in-control at $IG(42, 66)$ until observation 150, when it shifts to $IG(52, 66)$. The change is signaled at observation 193.

5.3 Conclusions

We have seen how two standard embellishments can be applied to the CUSUM scheme for processes well modeled by the inverse gaussian distribution. The fast initial response allows improved sensitivity to early departures from control, at the expense of slightly slower response to later departures. The self-starting CUSUM allows control of the process in the early stages, before enough historical data has been gathered to firmly establish the in-control parameters of the process.

Chapter 6

Bivariate Shewhart control charts

In this chapter, we generalize the idea of using highest probability density regions (HPD regions) to control two parameters of a process simultaneously. Doing so gives the tightest possible control limits. This can be useful in many contexts. We also address a vexing diagnosis problem found with traditional rectangular charting methods.

6.1 Current practices

The traditional control chart scheme maintains separate charts for each process parameter. For example, the charts for normally distributed processes include one for location and one for scale. If either chart signals, the process is deemed out of control. If the scale chart signals, the process is deemed to have undergone a scale shift. If the location chart shifts, however, it is not immediately clear whether that is due to a location shift or a scale shift. In that case, the accepted practice is to first examine the scale chart to see if there is any indication of a scale shift. In its absence, only then does one assume that there has been a location shift.

The operation of two charts as above is similar to running one bivariate chart with rectangular limits. See Figure 6.1. If a point plots inside the rectangular limits, then the process is assumed to be in control. If the process plots in region I (signaling a scale shift) the process is assumed out-of-control for scale. The process is only considered out of control for location if a sample is plotted in region II. See, for example, Montgomery [1991], who says, "Never attempt to interpret the $\bar{X}$ chart

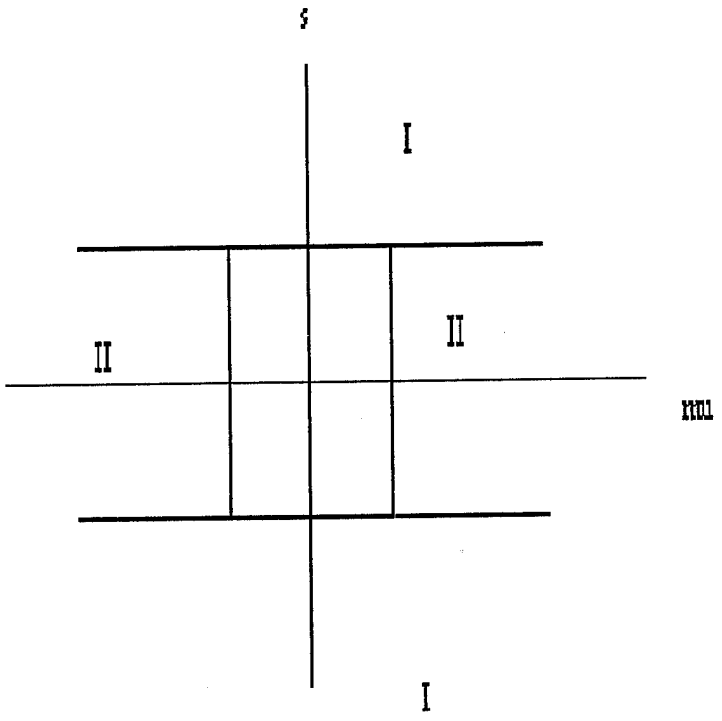


Figure 6.1: A bivariate Shewhart control chart with rectangular limits. Two possible out-of-control regions are labeled I and II.

when the R chart indicates an out-of-control condition". (An R chart is another type of control chart for scale, not discussed here.)

6.2 Improved diagnosis of an out-of-control signal

This algorithm can be improved by dividing the out-of-control region into three areas, instead of two. We will classify out-of-control points as either a location shift, a scale shift, or both.

Our approach will be based on likelihood ratios. Our first step is to hypothesize that, given an out of control signal, both the location and scale have shifted to their (now) most likely values, the MLEs. We then construct regions indicating a scale shift or a location shift only, based on a likelihood ratio with the parameters being either both of the MLEs (signaling both parameters have shifted) or just one of the MLEs (corresponding to only one parameter shifting). We examine

$$\Lambda_{location} = \frac{f(\overline{x}, s^2|\mu = \overline{x}, \sigma^2 = \sigma_0^2)}{f(\overline{x}, s^2|\mu = \overline{x}, \sigma^2 = s^2)} \tag{6.1}$$

$$\Lambda_{scale} = \frac{f(\overline{x}, s^2|\mu = \mu_0, \sigma^2 = s^2)}{f(\overline{x}, s^2|\mu = \overline{x}, \sigma^2 = s^2)} \tag{6.2}$$

If either Λ is greater than some critical value, we reject the hypothesis that both parameters have shifted in favor of the hypothesis that only one has shifted. This means that the likelihood of a shift of just one parameter is not much smaller than the likelihood of both parameters shifting.

For the normal case, we use charts for the mean and sample variance. Then

$$-2\ln\Lambda_{scale} = -n\ln(s^2) + 2n\ln(\sigma_0) + (n-1)(s^2 - \sigma_0^2)/\sigma_0^2 \tag{6.3}$$

which does not depend on $\overline{x}$. This implies that if we set $-2\ln\Lambda_{scale} < c$, we obtain boundaries parallel to the $\overline{x}$ axis.

On the other hand,

$$-2\ln\Lambda_{location} = \frac{n(\overline{x}-\mu)^2}{s^2} \tag{6.4}$$

which implies that accepting the hypothesis of a mean shift only depends on both $\overline{x}$ and s^2, and we obtain an acceptance region bounded by a quadratic curve.

The distribution of $-2\ln\Lambda_{location}$ is either central or non-central $F_{1,n-1}$, depending on whether the mean has shifted or not. The distribution of $-2\ln\Lambda_{scale}$ is not a standard one, and also depends on the correct specification of σ^2. Appealing to asymptotic theory for the distribution of $-2\ln\Lambda_{scale} \sim \chi^2_1$ [Bickel and Doksum, 1977],we obtain working critical values.

For the test of location shift against both location and scale shift, we use the appropriate critical value from the F distribution.

Note this .05 significance level is not for the test of whether the process has gone out-of-control; rather, it is for the discriminating between the hypotheses that both parameters have shifted or only one has shifted given that we already have a signal that the process is out-of-control. Loss considerations could motivate us to select other critical values, if there were higher costs associated with mis-diagnosis of one state.

For example, using a .05 significance level, $\mu_0 = 0$, $\sigma = 1$, and $n = 5$, we obtain a $\chi^2_{4,.05}$ critical value of 9.4877. Setting Equation 6.2 equal to the critical value, we obtain two roots: $s^2 = .07132$ and $s^2 = 5.5037$. Setting Equation 6.1 equal to f_{crit}, we obtain a parabola: $s^2 = \frac{5\overline{x}^2}{7.7086}$.

Further exploration of the exact critical values is deferred for future work. We note, however, that from the preceding paragraphs the shape of the regions is known. This, coupled with approximation theory and simulation, allows for the selection of reasonable approximations to the exact boundaries.

We use these regions to determine our diagnostics. Say our rectangular region

has been constructed to have an α significance level. Then, given we have an out-of-control signal, if $s^2 \in (.07132, 5.5037)$, we have a mean shift only. If $s^2 > .6486\overline{x}^2$, we have a scale shift only. Otherwise, we proceed under the conclusion that we have both a scale and a location shift. Figure 6.2 illustrates the regions.

The diagnostic boundaries are derived without regard to the control boundaries. Accordingly, it is possible to have an out-of-control signal which does not fall in the diagnostic regions. We shall see an example of this later. Accordingly, there is a fourth diagnostic state: indeterminate.

In summary, we are able to classify points out-of-control as arising from a mean shift only, a scale shift only, both location and scale, or unknown. Here we have used "location" and "scale" to represent the parameters controlled. The same process applies to the control of other parameters, such as shape.

6.3 HPD bivariate control regions

Rectangular limits are known not to give the smallest possible region for a given significance level. The HPD region gives that smallest possible region. However, the HPD region is not rectangular, and is therefore not the intuitive first choice when dealing with a pair of statistics for location and scale, especially when the statistics are independent when the process is in control.

We propose the following scheme. First, determine the HPD in-control or acceptance region for the joint density of the sampling distributions. This region will have the form

$$R_k = \{(x, y) | f_{ic}(x, y) \geq k\} \tag{6.5}$$

where $f_{ic}(x, y)$ is the in-control density. k is found numerically or by simulation for a given significance level.

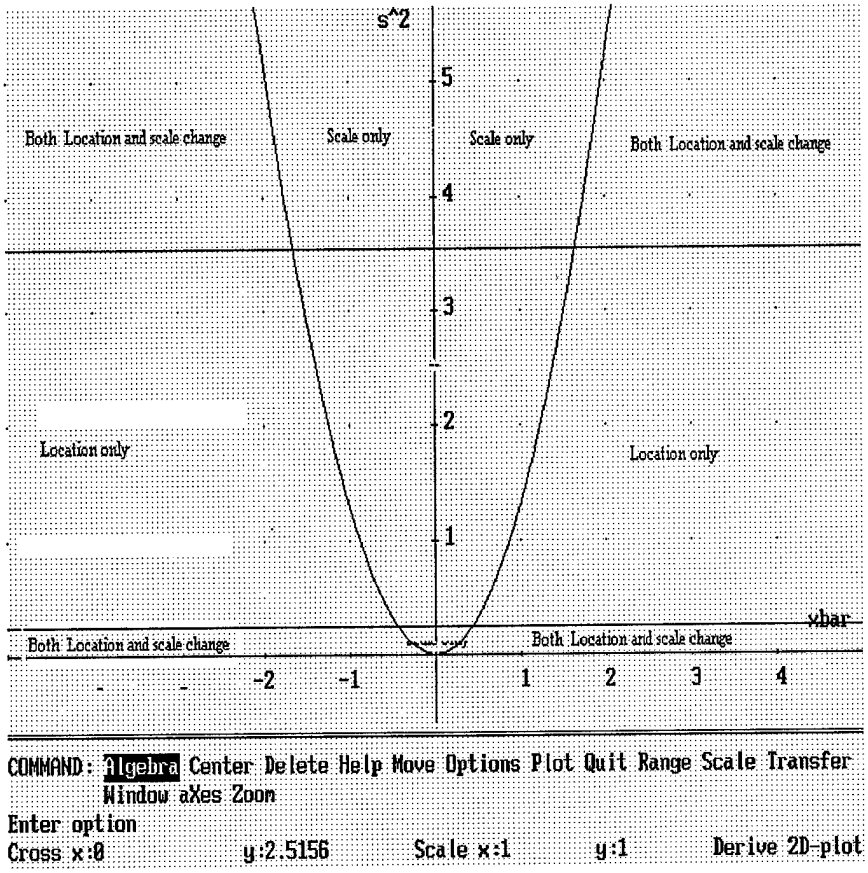


Figure 6.2: Diagnostic regions for bivariate Shewhart chart. Given a signal out-of-control, we classify the signal as either a location shift, scale shift, or both, depending where the signal is located. These boundaries are for a $N(0,1)$ in control. The rectangular control region is omitted.

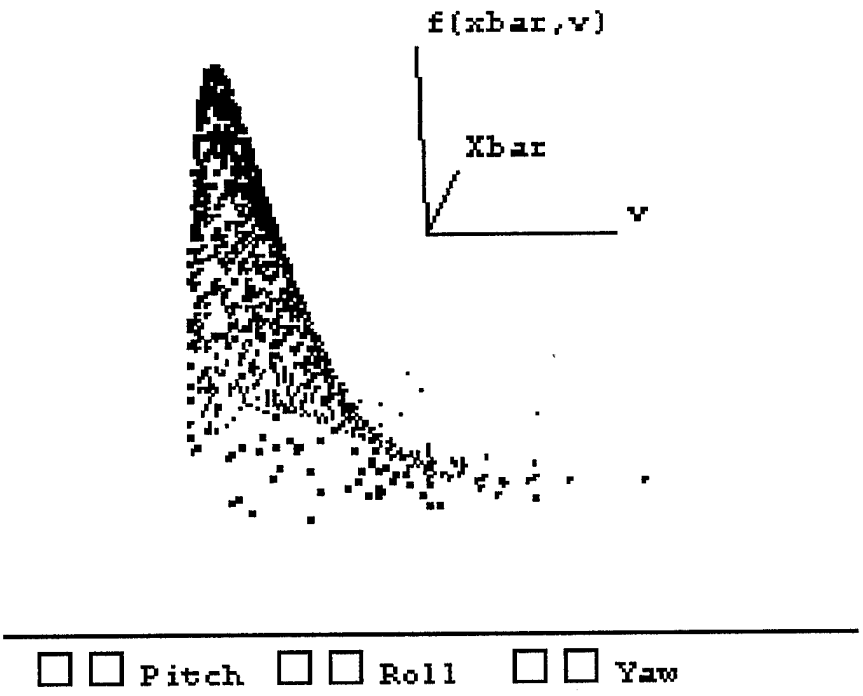


Figure 6.3: Spin plot of $\bar{X}$, V, and $f(\bar{X}, V)$ for 1000 samples of size five from an $IG(3, 5)$ distribution.

We propose this scheme for any bivariate distribution, but we will demonstrate it with an example using the inverse gaussian sampling distribution.

By illustration, we offer Figures 6.3, 6.4, and 6.5. Figure 6.3 is a three-dimensional spin plot of $\bar{X}$, V and $f(\bar{X}, V)$ for 1000 samples of size 5 from an $IG(3, 5)$ distribution. Figure 6.5 is a plot of the solution to $f(\bar{X}, V) = k$, with k selected so that $P(f(\bar{X}, V) < k) = .01$.

Second, plot the bivariate observations as they occur, signaling if the bivariate

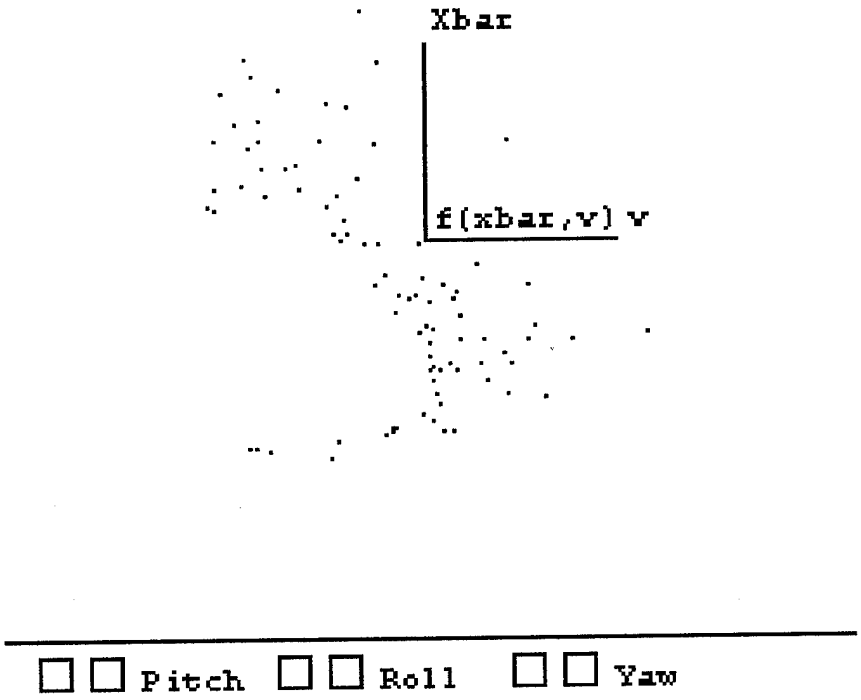


Figure 6.4: HPD out-of-control region by simulation. Spin plot of $\bar{X}$, V, and $f(\bar{X},\ V)$ for 1000 samples of size five from an $IG(3,5)$ distribution, censored to show the 100 points with the smallest values of $f(\bar{X}, V)$, rotated for effect.

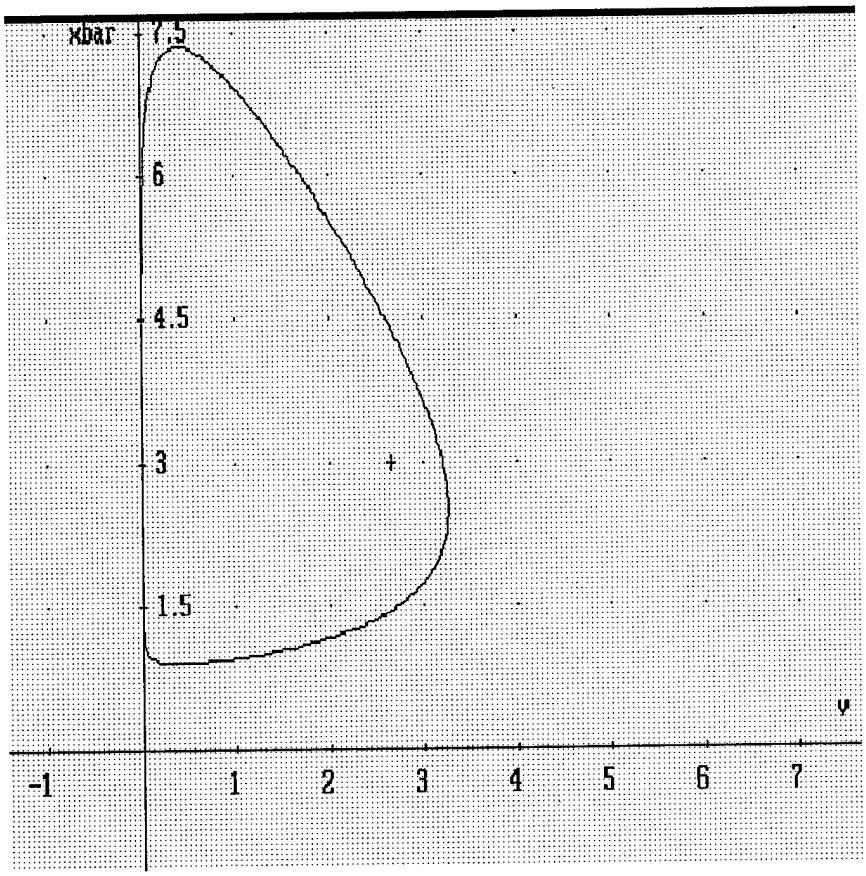


Figure 6.5: Graph of $f(\bar{X},\ V) = k$, where $P(f(\bar{X},\ V) > k) = .01$, for samples of size five from an $IG(3, 5)$ distribution. The exterior of the curve is the HPD rejection region. Compare with Figure 6.4.

observation is outside the HPD acceptance region. Interpretation of the signal will be discussed below. Third, to reduce chart clutter, use a weighting scheme that reduces the intensity of the plotted points on the (electronic) chart until points older than a given number of observations disappear. Make the most recent observation green; make any out of control observations red; make all others black.

Calculation of the bivariate HPD region requires solution of an integral equation: find k such that

$$\int_{R_k} f(x,y)dxdy = p \tag{6.6}$$

for $p = 1 - \alpha$. A solution to Equation 6.6 is found either by a numerical search method, based on Newton-Raphson, or by a routine to simulate the distribution of $f(X,Y)$ and determine the appropriate quantile. This approach is similar to the one used earlier to find the HPD region for the predictive schemes.

Calculation of the in-control ARL is not necessary: one simply inverts the significance level.

Calculation on the out-of-control ARL is easily (if slowly) accomplished using a numerical integration routine. Let $f_{ic}(x,y)$ be the joint sampling density, as before, when the process is in control, and $f_{oc}(x,y)$ the joint density when the process is out of control. Let $I_{k,f_{ic}(x,y)}(x,y)$ be the indicator function which is 1 if $f_{ic}(x,y) < k$. One finds:

$$1 - \beta = P(\text{signal}|\text{ out-of control}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{k,f_{ic}(x,y)}(x,y) f_{oc}(x,y)\ dx\ dy \tag{6.7}$$

One can find the ARL for any out-of-control state using Equation 6.7, since the out-of-control ARL will be

$$ARL = \frac{1}{1-\beta}$$

This approach for declaring out-of-control regions is applicable to any distribution. We shall provide examples using two distributions: the normal distribution and the inverse gaussian distribution.

6.3.1 An example

Here is an example. Let $X \sim N(\mu, \sigma^2)$. We take samples of size, say, 5. Normal practice would chart $\bar{X}$ and S^2. $\bar{X} \sim N(\mu, \frac{\sigma^2}{5})$ and $\frac{4S^2}{\sigma^2} \sim \chi^2_4$, and the two sample statistics are independent. Therefore, the joint density in-control of $(\bar{X}, S^2)$ is given by the product of their marginal densities. Now, in-control, let $\mu = 0$ and $\sigma = 1$. Then the joint density is:

$$f_{ic}(x, y) = \frac{2\sqrt{10}y \exp(-2y - 2.5x^2)}{\sqrt{\pi}} \tag{6.8}$$

with $x = \bar{x}$ and $y = s^2$.

We set the ARL at 100, giving $\alpha = .01$. Using an *Xlisp-Stat* routine available from the author, we determine k by simulation, obtaining $k = 0.004845$.

Now let the out-of-control distribution be given by $X \sim N(1, 1)$. The out-of-control ARL is given by Equation 6.7, which simplifies to $ARL = 3.98$.

Figure 6.6 illustrates the situation. Level curves for the in-control density are plotted, along with one out-of-control level curve.

6.4 Implementation of basic scheme

There are three computational issues when implementing this scheme.

First, finding the appropriate value of k requires solving a two dimensional integral equation involving indicator functions. This can only be done numerically, requires iteration, and is very slow, especially when using double precision arithmetic. We used simulation techniques for a fast approximation.

Second, plotting the control HPD limits requires an ability to plot implicitly defined functions of the form $f(x, y) = k$, which is not supported by all graphical packages, particularly *XLISP-STAT*.

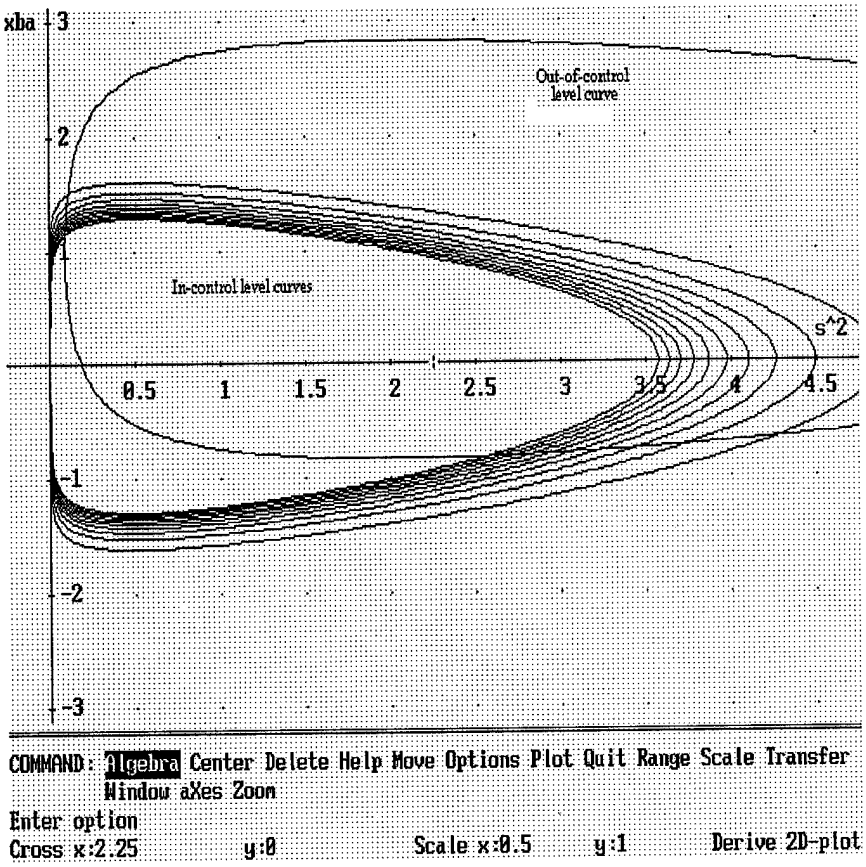


Figure 6.6: Level curves for the bivariate Shewhart chart normal example. The in-control curves are from a $N(0,1)$ distribution. The single out-of-control curve is from a $N(1,2^2)$ distribution.

Last, using the technique of varying colors and point intensity requires some fairly advanced programming skills.

We have a patchwork approach to implementation. First, we can find k using a FORTRAN program (again available from the author) that runs very slowly, or by simulation. Second, once we have k, we plot the control limits using *Derive*, which easily plots the implicit functions. Third, we do the actual plotting of points in *XLISP-STAT* which allows the coloring of points and varying of intensity. We were unable to find and unwilling to develop a routine to do implicit plots in *XLISP-STAT*.

With additional programming effort, the three tasks could be combined in one software application, but the patchwork is sufficient for demonstration purposes.

6.5 Comparison with traditional charts

We will compare the performance of the bivariate chart proposed here with that of the traditional Shewhart chart.

Both charts have the same in-control behavior, since both are constructed for the same significance level. When the k level for the HPD is found by simulation, the in-control ARL may not be exact. The ARL for the estimated k should be checked to assure ourselves that there is not a significant departure from the design ARL.

We compare charts by their out of control ARL for various combinations of out-of-control parameters.

6.5.1 Normal case

For our normal case, we will use $X \sim N(0, 1)$ as our in-control distribution. We take samples of size 5. We design for an ARL of 100.

Consider first our bivariate Shewhart method. Then our joint density was

given in Equation 6.8 for the case $n = 5$, $\mu = 0$, and $\sigma = 1$. We earlier found $k = .004845$.

For our various cases, we then compute (numerically or by simulation) our out-of-control ARL, by solving the integral equation in Equation 6.7.

We compare this with the standard normal case. For each of the two charts, we choose the significance level $\alpha_i = 1 - \sqrt{1-\alpha}$, so the rectangular region has the same significance as our bivariate chart.

Our control limits for $\bar{X}$ are given by $\mu \pm z_{crit}\sigma/\sqrt{n}$, which in this example reduces to

$$-2.8062/\sqrt{5} < \bar{X} < 2.8062/\sqrt{5}$$

Our control limits for S^2 are given by

$$\left(\frac{\chi^2_{1-\alpha/2,n-1}\sigma^2}{n-1} < S^2 < \frac{\chi^2_{\alpha/2,n-1}\sigma^2}{n-1}\right)$$

which in this example reduces to

$$.145053/4 < S^2 < 16.4183/4$$

Our out-of-control ARL for the standard charts in this example is found as follows:

$$p_{\bar{X}} = P\left(\left(-1.645/\sqrt{5} < \bar{X} < 1.645/\sqrt{5}\right) | X \sim N(\mu,\sigma)\right) \quad (6.9)$$

$$p_{S^2} = P\left(\left(\frac{.14503}{4} < S^2 < \frac{16.4183}{4}\right) | X \sim N(\mu,\sigma)\right) \quad (6.10)$$

$$ARL = \frac{1}{1 - p_{\bar{X}} \times p_{S^2}} \quad (6.11)$$

We tabulate some of these values for various out-of-control cases in Table 6.1. We obtained the value of k for the bivariate case by simulation, which explains why the ARL is not exactly 100 for the in-control case.

μ	σ	ARL–bivariate	ARL–standard
0	1	99.20	100.00
1	1	3.98	3.47
2	1	1.07	1.05
.5	1	21.97	19.79
0	1.5	4.22	5.75
0	.5	8887.	28.76
1	1.5	2.00	2.32
1	.5	11.9914	6.3622

Table 6.1: Comparison of ARLs for bivariate Shewhart and a pair of standard charts. ARLs are given in number of samples of size 5. The in-control distribution is $N(0,1)$.

We see from the Table 6.1 that we have improved power for increases in σ^2, reduced power for decreases in σ^2, and comparable power for detection of shifts in location. We also see that we have improvement in our detection of simultaneous shifts in μ and increases in σ.

As an example of the charting, we take samples of size 5, with the process in-control as $X \sim N(0,1)$. It shifts at observation 10 to $X_o \sim N(1,1)$. We show the charts for observations 9 onward in Figures 6.7 through 6.13. For this black and white printer, the points in control are marked by "·"s of various sizes, with larger points being more recent. Also, due to difficulty implementing the plotting implicit functions, the control limit curve is not shown. The points out of control are marked by red "◇"s. The most recent point is given by "+", if in-control. To avoid clutter, only the 9 most recent in-control points are displayed. Out-of-control points remain visible indefinitely.

When the process goes out of control, we are able to diagnose the out-of-control point using the rules portrayed in Figure 6.2 to see that we most likely have a mean shift only.

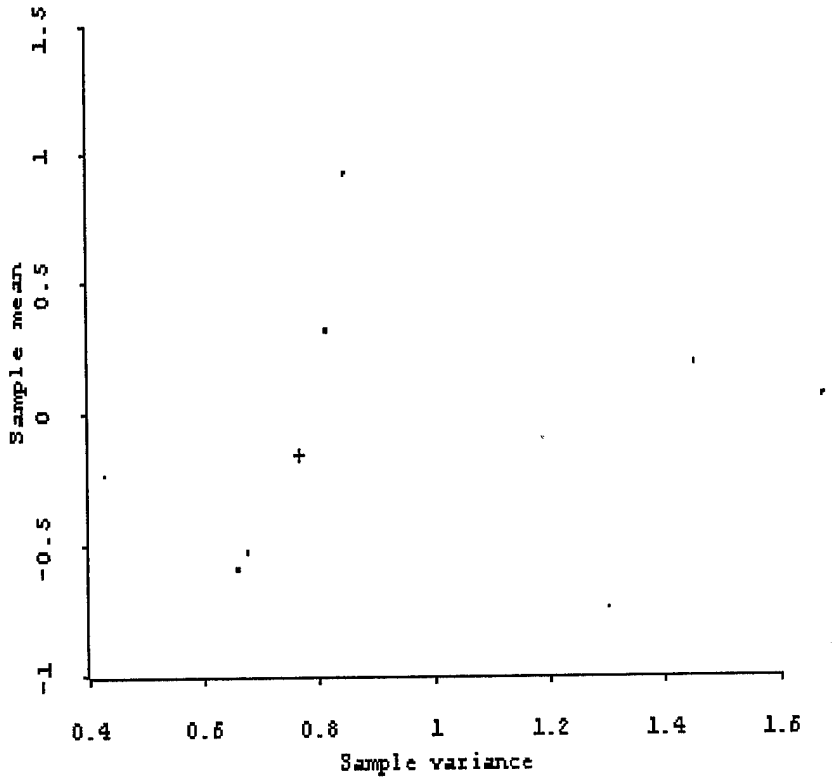


Figure 6.7: Bivariate Control chart for Normal Example, up to observation 9.

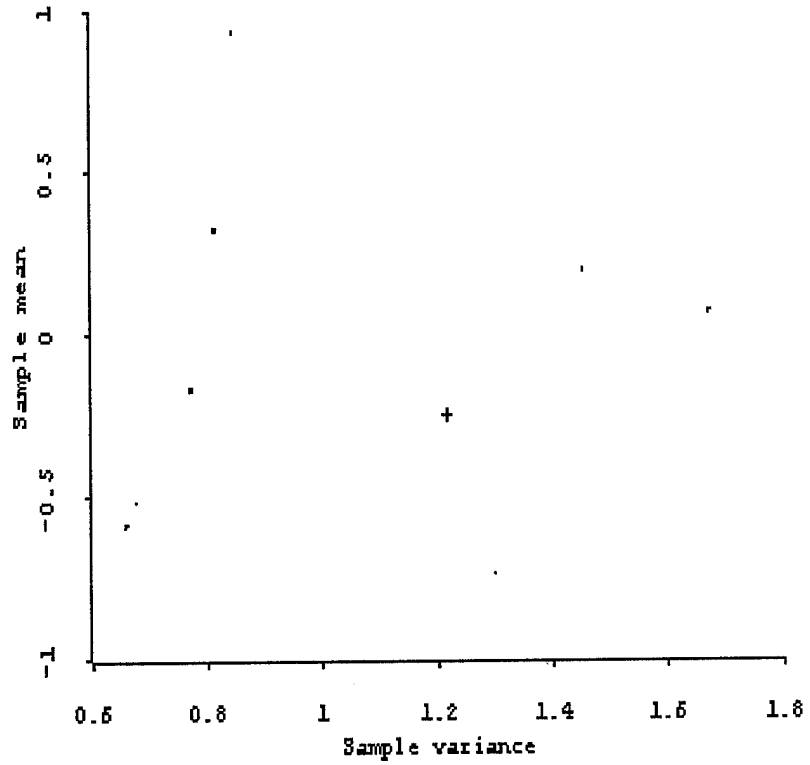


Figure 6.8: Bivariate Control chart for Normal Example, up to observation 10. Notice only the most recent 9 points plot.

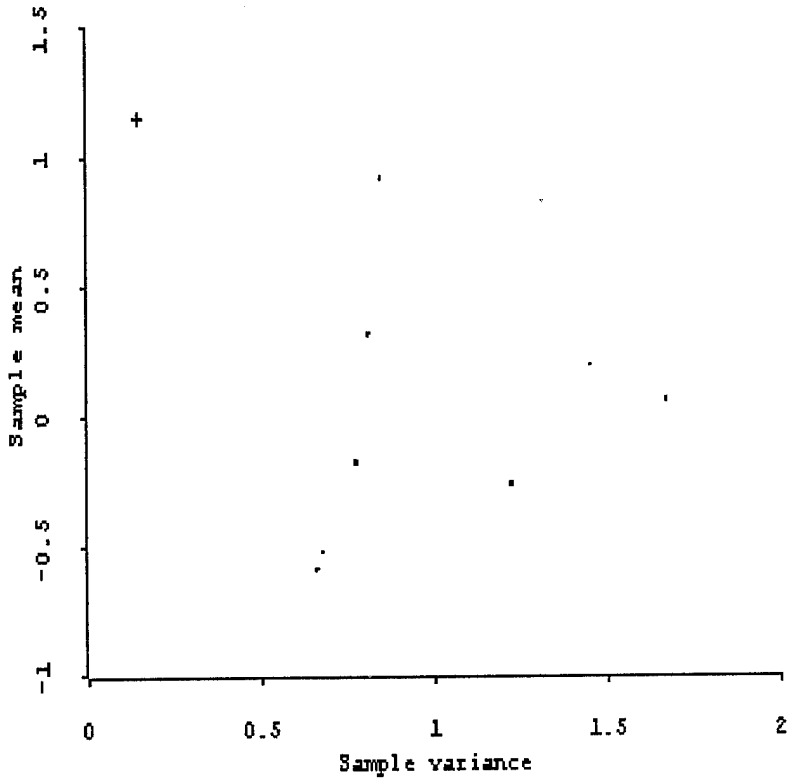


Figure 6.9: Bivariate Control chart for Normal Example, up to observation 11. The process is now out of control at $N(1,1)$. The chart has not yet signaled.

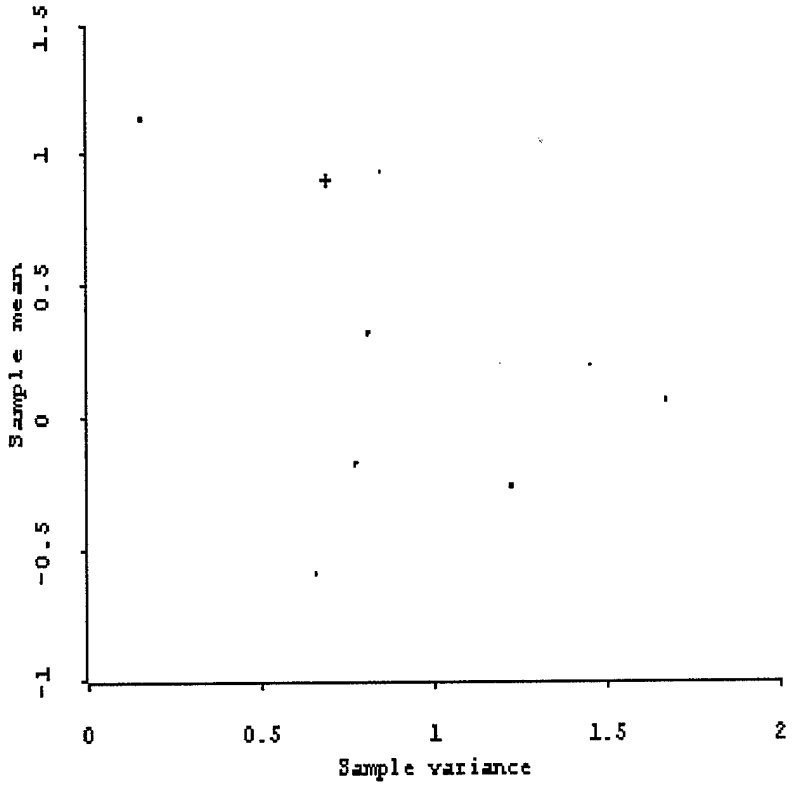


Figure 6.10: Bivariate Control chart for Normal Example, up to observation 12. The process is now out of control at $N(1,1)$. The chart has not yet signaled.

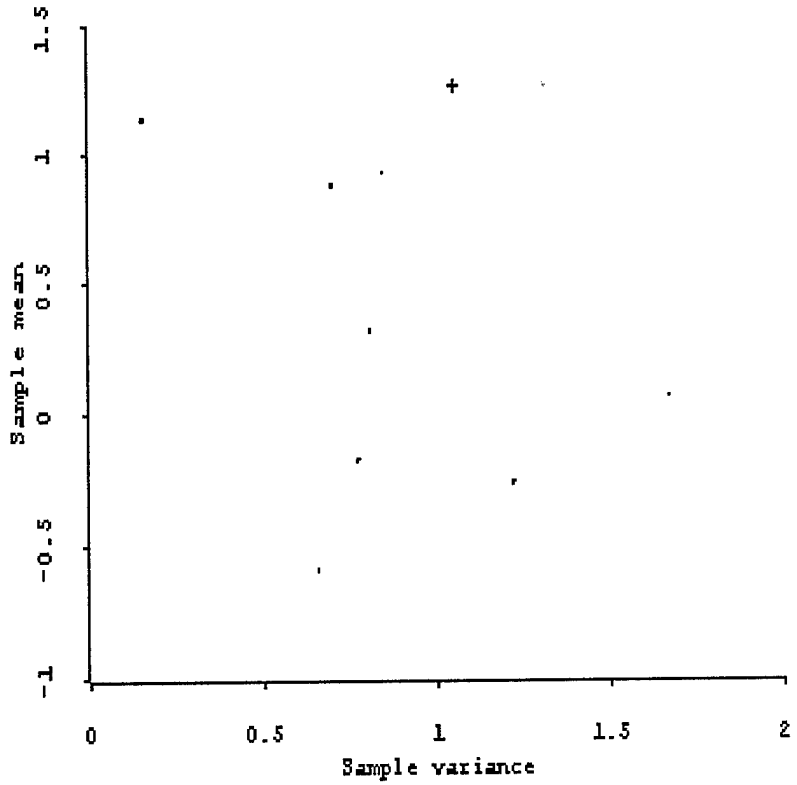


Figure 6.11: Bivariate Control chart for Normal Example, up to observation 13. The process is now out of control at $N(1,1)$. The chart has not yet signaled.

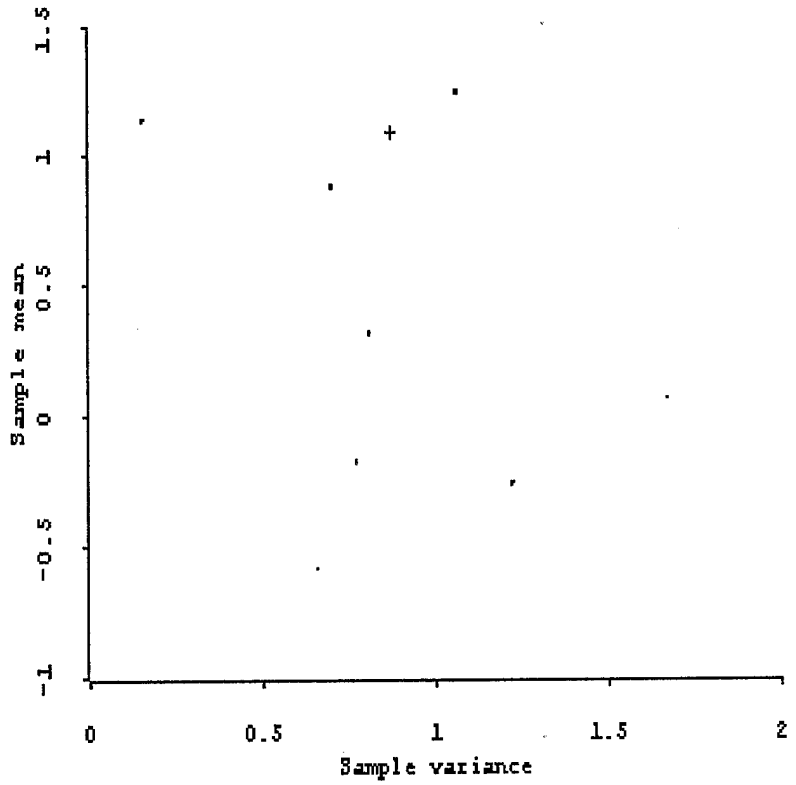


Figure 6.12: Bivariate Control chart for Normal Example, up to observation 14. The process is now out of control at $N(1,1)$. The chart has not yet signaled.

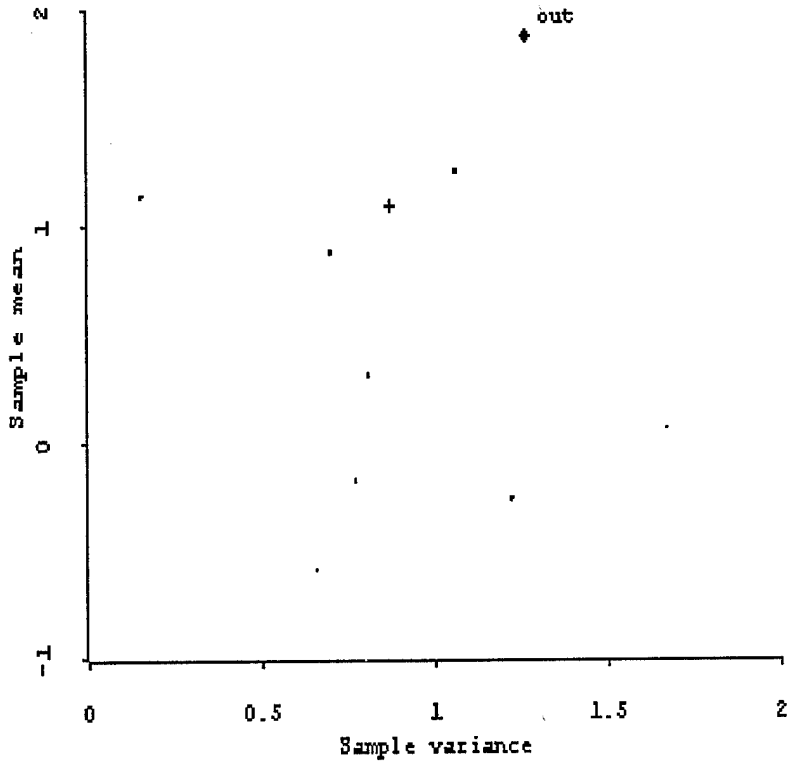


Figure 6.13: Bivariate Control chart for Example, up to observation 15. The process is out of control, and has now signaled out-of-control. Applying the rules developed in Figure 6.2, we diagnose a mean shift only.

6.5.2 IG case

We now turn to the IG case, and develop bivariate Shewhart charts based on the HPD for this distribution.

The joint density for $\overline{X}, V$ is given by

$$\begin{aligned} f(x,v) &= \sqrt{\frac{n\lambda}{2\pi}}\overline{x}^{-3/2}\exp\left(-\frac{n\lambda(\overline{x}-\mu)^2}{2\mu^2\overline{x}}\right) \\ &\quad \times \frac{\lambda}{2^{(n-1)/2}\Gamma((n-1)/2)}(\lambda v)^{(n-1)/2-1}\exp\left(-\lambda v/2\right) \end{aligned} \tag{6.12}$$

We will define our control region by using the highest probability density. We define $R_k = \{(\overline{x}, v) | f(\overline{x}, v) \geq k\}$ and set k so that $ARL = \frac{1}{P((\overline{X},V)\in R_k)}$ for our desired ARL. As with the normal case, this requires finding a solution to Equation 6.6. This is no less difficult than in the normal case. We approximate k by simulation.

We use a similar diagnostic scheme, based on likelihood. Using $-2\ln\Lambda$, we obtain diagnostic regions for V and $\overline{X}$ when the process signals out of control.

If $2\ln(n-1)^{n/2} - n\ln(\lambda v) + \lambda v - n + 1 \leq c$, we declare the process out of control for a scale shift. This equation will have two roots, so we obtain a region of the form $V < c_1$ or $V > c_2$.

If $\frac{n(n-1)(\overline{x}-\mu)^2}{\mu^2 v\overline{x}} \leq d$, we declare the process out of control for a mean shift only. This equation will be a quadratic, and is similar in form to the equation we obtained for the normal diagnostic curve.

As before, if we don't signal an exclusive shift, we assume that both parameters have shifted.

6.5.3 An example

We assume that we have an $IG(3,5)$ process in-control. We draw samples of size 5. We desire an ARL of 100. By simulation, we determine that $k = 3.0\ \ 10^{-4}$. Our HPD and diagnostic lines are plotted in Figure 6.14. We run the bivariate chart for

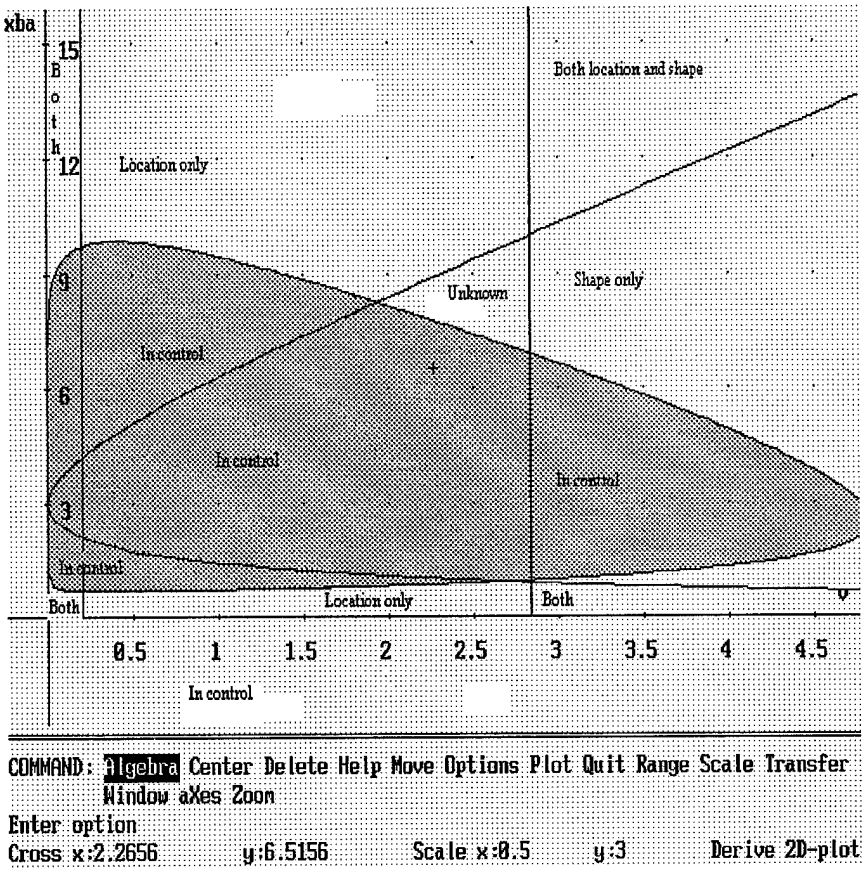


Figure 6.14: Bivariate HPD regions for an $IG(3,5)$ with samples of size 5. The in-control region is shaded. Out of control areas are labeled with their diagnosis. Note the scales.

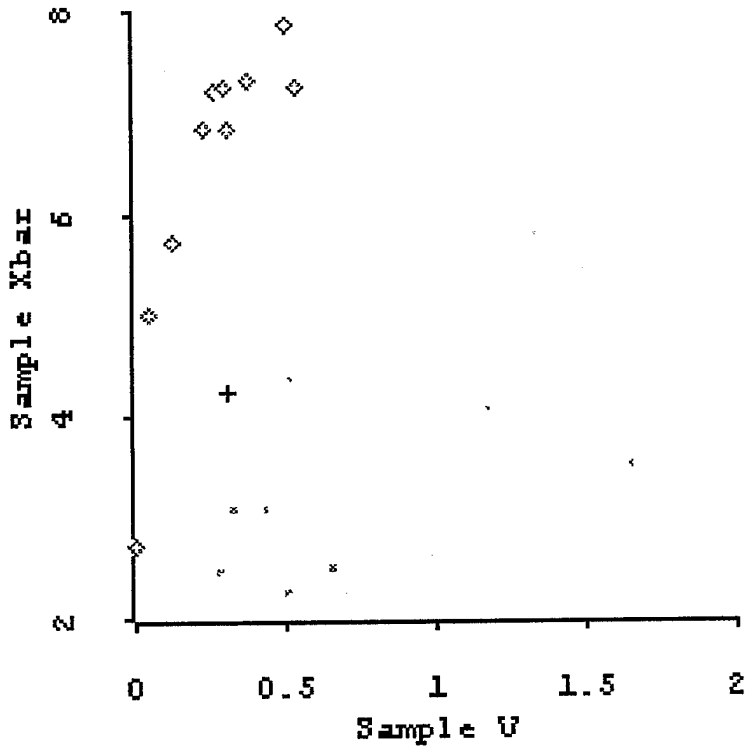


Figure 6.15: Long run Bivariate HPD chart for $IG(3,5)$ with 1000 observation. Only the last 9 in-control points are plotted. There are 10 outliers, marked with diamonds.

1000 observations in control, plotting as before the last 9 observations plus all out-of-control observations. We observe 10 points out-of-control, illustrated in Figure 6.15. This is exact agreement between observed and predicted number of out-of-control observations.

We shift the process to an $IG(6,4)$ and continue to take samples of size 5. In our 1000 points, we observe 299 observations signaling out-of-control. The chart for this run is shown in Figure 6.16.

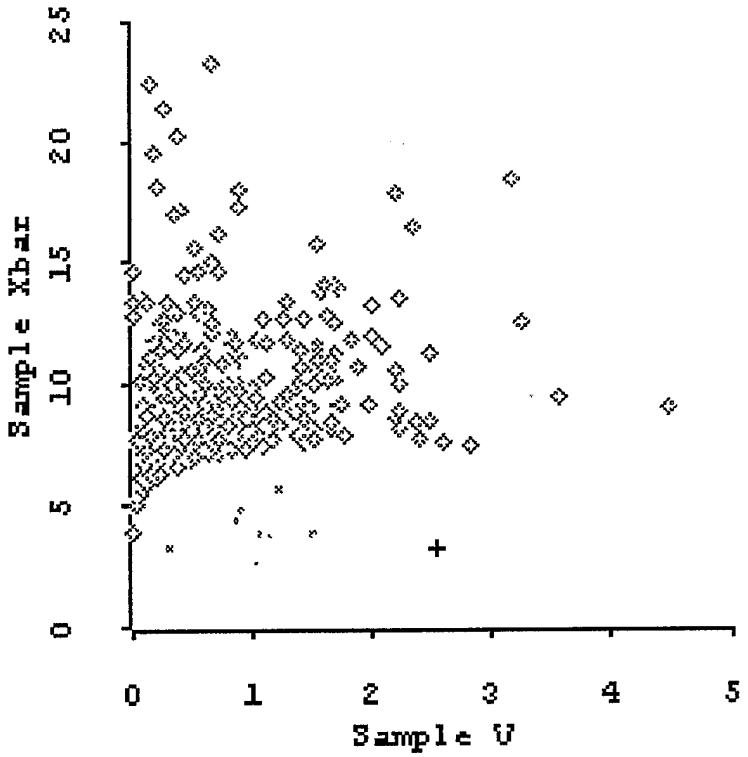


Figure 6.16: Out-of-control bivariate HPD chart, with 1000 observations.

6.6 Interpretation of a signal

When an observation is out of control, we can automate the diagnosis by including the classification algorithm in the computing code. When the process signals, the point is plotted in an out-of-control region, which visually indicates to the operator an initial diagnosis. We can also print the out-of-control values, and report the likelihood ratios for a mean shift only against a shift in both mean and scale, and a scale shift only against a shift in both.

With any diagnostics based on likelihood, it is possible for an observation out of control for one reason to plot in a different region. Operators should be especially alert to this possibility when the out-of-control observation is close to a boundary.

6.7 Conclusions

We have proposed and developed two ideas in this chapter. First, we have advocated using bivariate control charts to control processes to allow better diagnosis of out-of-control signals. We applied this method to both the normal and inverse gaussian distribution. Second, we have examined using the HPD region as the control region for these bivariate charts.

In many contexts, one follows the same procedures to react to an out-of-control process regardless of its suspected cause. In other contexts, one proceeds quite differently based on the initial diagnosis. In those contexts, the improved diagnostic tools of this chapter save time and money by directing the corrective actions first to the most likely cause.

Chapter 7

Application to combat models

In the next two chapters, we apply the tools we have developed to problems of interest. In this chapter, we examine control of software revisions. In the next chapter, we look at an automobile assembly line.

7.1 Background

There are two major approaches to modeling combat. The first, originated by Frederick Lanchester [1956] at the turn of the century, represents combat by differential equations. The second, currently popular, involves high detail computer simulations. Each suffers from weaknesses. In this section, we propose a hybrid model based on the inverse gaussian distribution, which captures some of the advantages of both.

Given this hybrid model, we can monitor simulated or actual operations to detect model changes, using the tools developed in the preceding chapters.

Many authors have attempted to model combat. The first and arguably most influential was Frederick Lanchester. He proposed simple differential equation models for attrition, where the rate of change of the force level of one side was a function of the friendly force level and the opposing force level. The form of the function depended on the type of combat. The solutions to these differential equations have been used extensively in military modeling [Taylor, 1981] [Taylor, 1983].

These Lanchester Equations are recognized to have several shortcomings [Hughes, 1964] [Dupuy, 1987] [Ventisel, 1964]. They are deterministic, simplistic, and do not

fit the historical record well. Still, they are widely used because they are easily understood and may give insights despite their weaknesses. The Corps level model *Vector–in–Commander* used by the Army is a deterministic model based on Lanchester Equations.

A second approach has arisen with the advent of powerful computers. This is the high resolution simulation. In these models, every actor on the battlefield is modeled. Then a stochastic simulation is run, and the results reported. For example, one soldier may be set in motion towards an object. If he makes contact with an enemy soldier, there is a conflict resolution according to some stochastic algorithm. And the game proceeds according to the results. A large number of actors and a large number of conflicts produce a very large space of possible outcomes of the simulation.

Some of these simulations are interactive, with humans making decisions at appropriate points. Others are not; they run as programmed until some stopping criterion is met.

Unfortunately, there are serious questions about the utility of these large scale simulations. [Dupuy, 1987]

First, they are only as good as the underlying algorithms, and in many cases the algorithms rely on either Lanchester Equations (which are known to be flawed) or on Monte-Carlo models with the parameters estimated on an ad hoc basis. These assumptions are usually invisible to the user of the simulation, and tend to become obscured and lost even to those who are responsible to maintain and improve the simulations. This is not due to a lack of diligence or professionalism, but is a result of the sheer size of the programs, the turn-over of personnel, and the fundamental lack of good algorithms.

Second, these models also do not have good records replicating the historical record.

A third approach to modeling combat has been taken by those who have

attempted to construct statistical models based on the historical record. Robert L. Helmbold offers a survey of these results [Helmbold, 1990]. Despite the best efforts of many talented people, these statistical models have failed to explain the variation in the overall historical record well, with the best regression models enjoying $R^2 \approx .3$.

These issues are not just of interest to the military academic. Procurement, doctrine, and force structure decisions are being made on the basis of the results of military models. This is a trend that will continue and accelerate, as the cost of conducting analysis based on physical models is prohibitive. Sound stewardship of national resources as well as prudence in the conduct of the national defense make it imperative that the models used be correct as possible.

7.1.1 Underlying hypothesis of Brownian motion for combat models

Combat operations are inherently stochastic. This nature argues in favor of deeper models than the Lanchester differential equations, which at best can be considered models of the expected results of combat and completely fail to capture the distribution of results that may occur.

It is possible to cast these equations as stochastic differential equations (SDEs), and attempt to solve them. However, complicated situations with many mixtures of players argue that such stochastic differential equations would be very complicated to solve. Additionally, the coefficients of such SDEs would need to be fit or estimated from either historical data or some other modeling effort.

In the remainder of this chapter, we make a large assumption. Without attempting to explicitly define the SDEs, their coefficients, or their number and type, we will assume that their solution is well modeled by Brownian motion with drift.

This is not such a huge assumption as it might appear. First, on its face it

appears a reasonable model of the physical nature of combat operations. Many individual actions occur as a force moves toward an objective, some favorable, and others unfavorable. The sum of these actions constitutes the net effort of the larger unit. It is not unreasonable to model this ebb and flow of battle as Brownian motion with drift. This seems especially unobjectionable when the characteristic we are modeling is actual movement of a unit. Attrition models require a stronger assumption, since attrition tends to accelerate as one side gains ascendancy.

Second, this type of model captures the benefits of the simplicity of the differential equation approach while retaining the distributional nature of the simulation approach. It also is much less computationally intensive than the simulation effort.

Third, it appears partially supported by the historical record. Helmbold [1990] looked at advance rates for 634 battles. He concluded:

> The upshot of our analysis is the advance rates are *not* normally distributed. Their distribution is highly skewed, and much more closely fit by lognormal distributions than by any of the others tried (normal, exponential, Weibull, and gamma.)[Helmbold, 1990]

Helmbold did not try the inverse gaussian distribution. We know that data well fit by the log-normal distribution is usually also well fit by the inverse gaussian.

We examine the log-normal probability plot in Figure 3-1 of Helmbold. This is a plot of the advance rates observed in 57 battles in the Italian Theater during W.W.II between the fall of 1942 and spring of 1944. We notice that the data does appear to be well fit by the log-normal except at the tails, where there is a slight S shape. The exact same tail behavior is demonstrated by the graph of 1000 $\ln(IG(3,5))$ variates when similarly plotted in Figure 7.1. In other words, it is very plausible that the data from Helmbold is better fit by the inverse gaussian model than by the log-normal model. Without access to the original data, this is as compelling an argument as can

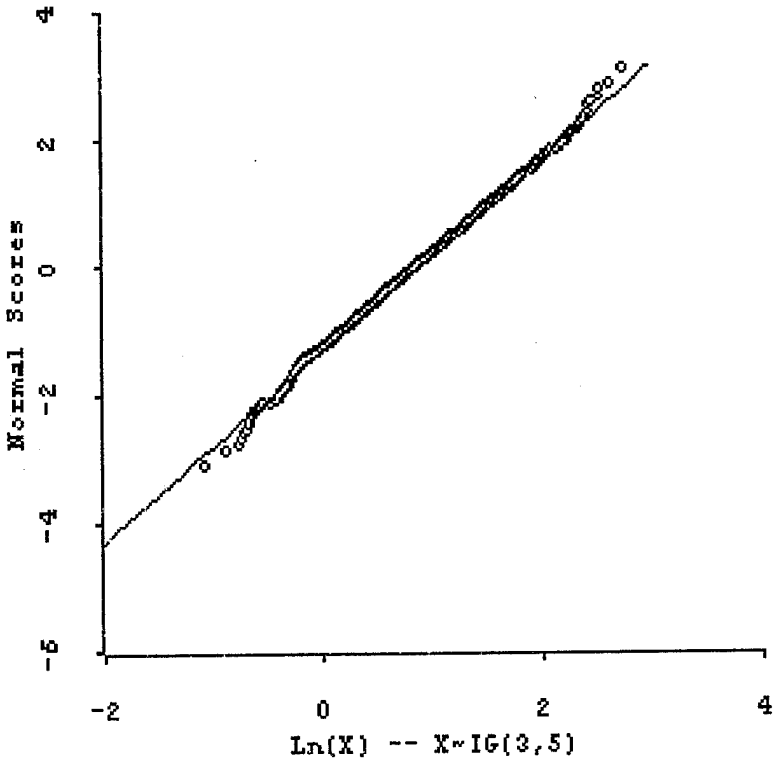


Figure 7.1: Example of the tail behavior when an IG random variable is plotted on a log-normal probability plot.

be made from the graphical evidence available. We are encouraged to continue our efforts to model advance rates with the inverse gaussian distribution.

Of course, what we would expect from the physical process is that the time to accomplish a mission was distributed as an inverse gaussian random variable. This too seems plausible. If the advance rate is well modeled as inverse gaussian, then the time to move a fixed distance is distributed as the reciprocal of an inverse gaussian. Since the distribution of the reciprocal of an inverse gaussian random variable is known and similar in form to an inverse gaussian, we proceed to model the movement itself as having an inverse gaussian distribution. We note that if this was true, then the log-normal probability plot of the advance rate would appear as in Figure 7.2, which again matches the figure in Helmbold and is similar to Figure 7.1.

A separate issue raises itself, and we defer it for future study. If the movement of forces appears to follow an inverse gaussian distribution, perhaps regression methods of explaining combat based on the inverse gaussian distribution, instead of the normal distribution, may be successful for statistical modeling of combat.

7.1.2 Program maintenance and unintended effects

The computer simulations used to model combat are very large programs. The programs are constantly being maintained, as new algorithms are added, new weapons systems are modeled, and new scenarios imagined. Historical data also causes model changes and updates, as when the experience of the Gulf War did not match the predictions of the simulations run during the planning phases.

The major Army proponent for these models is located at the Training and Doctrine Command Analysis Center (TRAC) activity at the White Sands Missile Range (WSMR). There, analysts, modelers, and programmers continually change the program code. To test for unintended programming effects, TRAC-WSMR has a

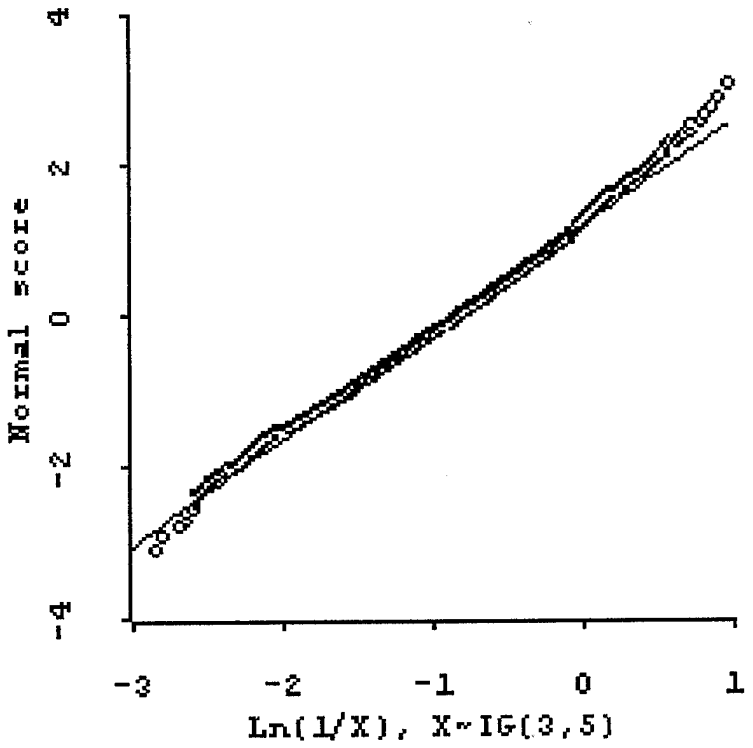


Figure 7.2: Example of the tail behavior when the reciprocal of an IG random variable is plotted on a log-normal probability plot.

suite of scenarios it runs after updating the programs to check for consistency with earlier versions of the software.

There are many measures associated with each model run. For example, one can measure friendly losses, enemy losses, the ratio of friendly to enemy losses, and so on. If we represent the number of friendly forces by X and the number of enemy forces by Y, then $\frac{\Delta Y}{\Delta X}$ is the *Loss Exchange Ratio.* This measure is widely used, and the one we examine in the remainder of this chapter.

We defer examination of other measures for later work for two reasons. First, additional post-processing of simulation runs is expensive, and we are already indebted to the Rand Corporation and to TRAC-White Sands Missile Range for the extensive work they did to produce data on the loss exchange ratios. Second, the loss exchange ratio is a widely used measure for modelers, and we shall see below that it appears to be well-modeled by the inverse gaussian distribution. This lends credibility and practical significance to the work, which would not be obtained by working with a less familiar measure.

We have proposed to TRAC-WSMR and to the Army Research Laboratory that the time to accomplish the mission and the time to make decisions be examined to determine if the inverse gaussian distribution is an appropriate model. The Army Research Laboratory has agreed to fund the author's investigation of those topics. That research will be conducted in the summer of 1996 in conjunction with a large command and control exercise at Fort Leavenworth, Kansas. We feel that those measures have a better theoretical basis for being modeled as inverse gaussian variates.

7.2 Goodness of fit

We have as our initial data 80 replications of a simulation conducted at Fort Hood, Texas, in December 1993. These trials were used to measure the effectiveness of the

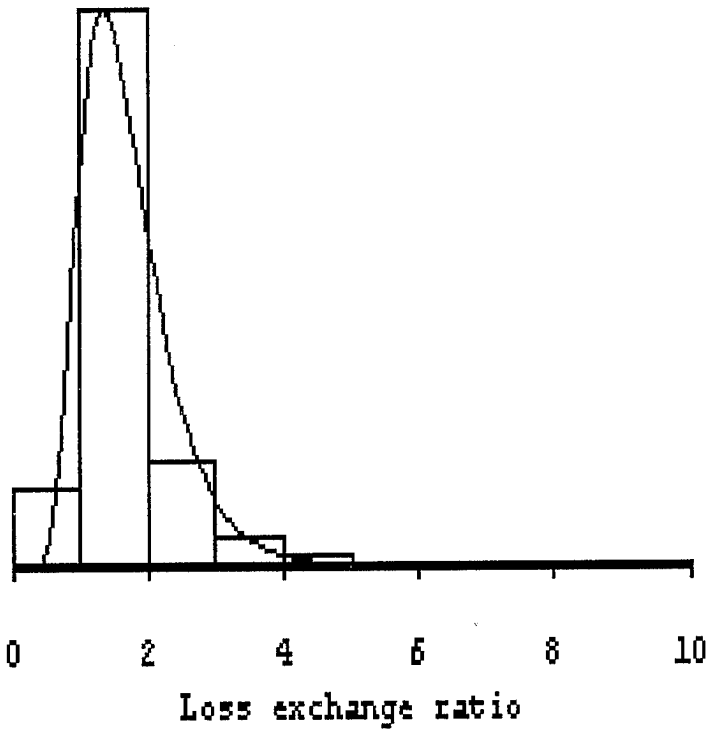


Figure 7.3: A histogram of force exchange ratios from the Blue Defense vignette for the M1A2 IOTE. The fitted IG distribution has been superimposed.

M1A2 main battle tank. The trials consisted of a Blue armor task force defending against a Red force. We were provided the friendly losses ("Blue losses"), the enemy losses ("Red losses"), and the resulting loss exchange ratio.

This model represents damage to each vehicle as a vector, representing of damage to different sub-systems. The model changes the allowed behavior of each vehicle, based on its damage vector. Different ways of representing this allowed behavior have been coded into the model, and simulated. These different approaches constitute our possible out-of-control observations.

The MLEs for the parameters for an inverse gaussian model of the base data are $\hat{\mu} = 1.697875$ and $\hat{\lambda} = 11.2150895$. A histogram of the data with the fitted density is provided at Figure 7.3 and indicates a good fit. We pursue goodness of fit testing next.

7.2.1 Goodness of fit of IG model

Following Edgeman, Scott and Pavur [1988], we perform a modified Kolmogorov-Smirnov test for the goodness of fit of the inverse gaussian distribution. The Kolmogorov-Smirnov statistic $D_N = 0.0982$. We further adjust this value using the regression equation $D_N^* = D_N(\sqrt{N} + b_1 N^{-.5} + b_2 N^{-1} + b_3 N^{-2})$, obtaining $D_N^* = 1.2753$.

The critical value for rejecting the hypothesis that the data is well fit by an inverse gaussian distribution at the .20 significance level is greater than 1.994, using Table I in Edgeman, Scott and Pavur. The critical value at the .10 significance level is at least 2.356. Accordingly, we find no evidence that the data is not well fit by an inverse gaussian distribution.

7.3 Results

We will now use the procedures in this thesis to detect changes to the underlying model. We have three possible out-of-control scenarios.

The first one represents a large model change. We will attempt to detect these changes with our self-starting and predictive Shewhart charts.

The second two scenarios correspond to small persistent model changes. These are the ones least likely to be noticed after program maintenance. For these we will use self-starting CUSUM charts.

We could, of course, apply all methods to all cases, but we have chosen to

limit ourselves.

We assume the base case corresponds to the process in-control. Since we have less than 100 in-control individual observations, we will use self-starting and predictive methods.

The first case we examine is a previous version of this computer combat model. The old methodology assigns a one-dimensional utility number to a combat vehicle. This number represents the fraction of capability left in the vehicle. It does not distinguish between loss of capability due to loss of mobility or that due to loss of a weapon system. We expect this model to perform differently from the base case. We have 20 observations for this case.

The second two cases are modifications to the base case. The second case allows only two states for each subsystem: operable or inoperable. The third case allows the utility for each subsystem to be any value in the interval $[0, 1]$. This contrasts with the base case, which has a finite number or degraded utility states. We have 80 observations of each of these cases.

Figure 7.4 shows a boxplot of the four data sets, with the base case at the left, then the old program, and the two further modifications.

Each of the four models tracks damage to the vehicles differently, and feeds that information into the main battle simulation. We note from Figure 7.4 that the old model is obviously different from the other three, but that the two modifications do not look very different from the base case.

Normally, we would not use control charts in this scenario, but rather tests of equality of parameters. This is especially true because we know precisely when the possible model changes occur. We use these four cases to illustrate a more sophisticated process. Over the life of this code, there are literally hundreds of changes in the details of the implementation. The four described above are not very significant programming changes, and refer to one very small piece of the code. How does one

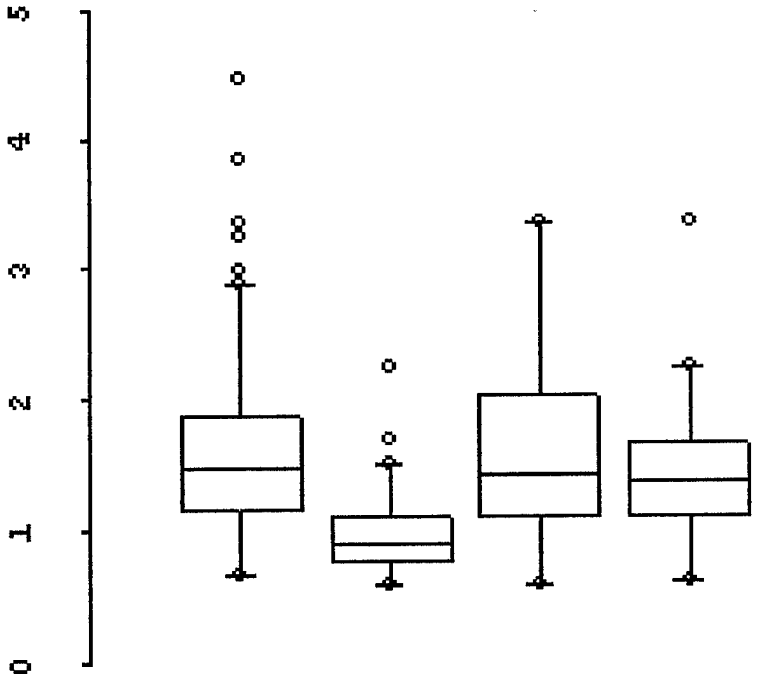


Figure 7.4: Boxplot of the four data sets used in this chapter. From left to right, they are the base case (assumed in-control), the old program, and two modifications to the base case. Source: White Sands Missile Range, 1996.

monitor the overall performance of the code when there are continual changes, most of which are not supposed to produce any change in the output characteristics? In this setting, control charts are useful to track the model over time, when it is unknown when a substantial change to the model may occur.

There is no history of statistical testing of equivalency of models after programming changes at WSMR, according to the sources who provided this data. We hope that this methodology will be adopted by WSMR, and we have reason to believe that it will be.

The concept of charting computer program performance against a set of benchmarks during program modifications is useful in more contexts than just this one military application. It allows a holistic view of software maintenance over time, in any context.

7.3.1 Self starting Shewhart Charts

Here we compare the base case with the old model, using a self-starting Shewhart scheme for the mean. The chart is at Figure 7.5. We see that even with this relatively large model change, we do not get a signal in 20 observations, although there is an obvious downward trend.

If we break the data into samples of size five, there is also no indication of a shape change on the self-starting Shewhart chart for shape, found in Figure 7.3.1. Perhaps the lack of evidence is due to the process not running long enough. To check this, we sample from the in-control data 200 times to produce a new "base case". We then check the self-starting charts for this data, again with just our original 20 out-of-control points.

We obtain an immediate signal on our self-starting Shewhart chart for μ with this strategy, as seen in Figure 7.3.1. We obtain no signal in our self-starting chart

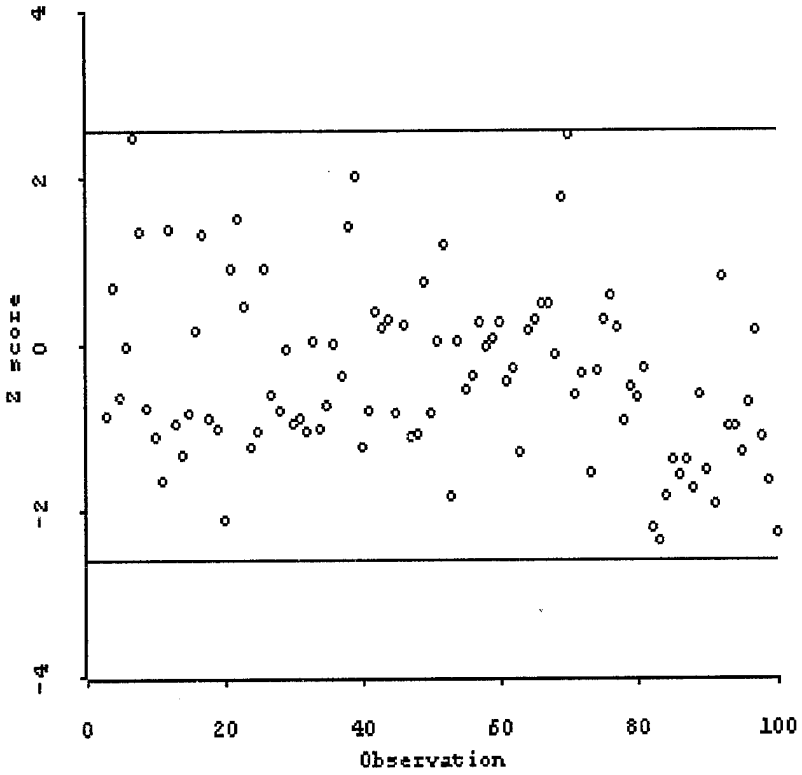


Figure 7.5: A self-starting Shewhart chart for the mean of the base case and historical data.

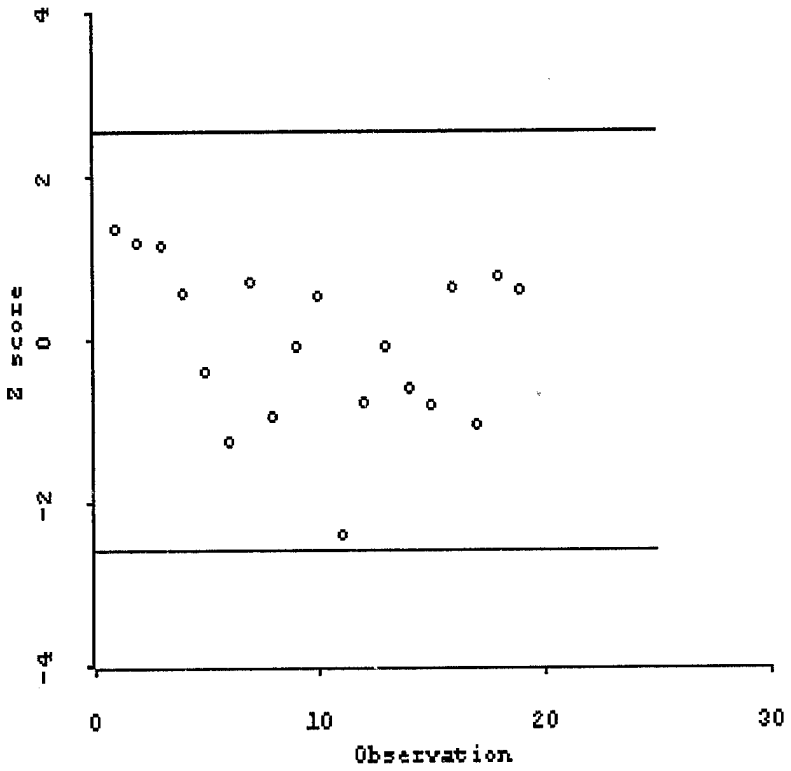


Figure 7.6: Self-starting Shewhart chart for λ. Data is treated as samples of size 5. No evidence of a shape change is found.

for λ.

We see no conclusive evidence in our original self-starting Shewhart charts of a model shift in either μ or λ. We do see evidence of a shift in the mean in our "bootstrap" self-starting Shewhart chart for the mean.

7.3.2 Predictive Shewhart Charts

We construct the predictive chart for the base case followed by the large shift in Figure 7.8. Just as the self-starting chart, the predictive chart does not signal, although it does indicate a downward trend in the data.

We increase the size of the training set by bootstrapping the base case to 200 points, and construct at Figure 7.9 the predictive chart for the data. We see that, unlike the self-starting chart, the predictive chart does not signal immediately. There are several interesting features to this chart. First, there are 7 out-of-control points in the first 200 observations. With an ARL of 100, we expect two out-of-control points. This indicates that the process may not be in control. Since these points are normalized from the predictive scheme, that indicates that the process may not be modeled well by the inverse-gaussian. This is interesting, and argues for more precise distributional testing of the White Sands data.

7.4 Self-starting CUSUM charts for the mean

We apply the self-starting CUSUM schemes to the comparison of the base case with all three model changes. The charts are displayed in Figures 7.10, 7.11, and 7.12.

We see that the self-starting charts detect the mean shift quickly for the shift from the base case to the historical case, and relatively quickly for the shift from the base case to the other two degraded states models.

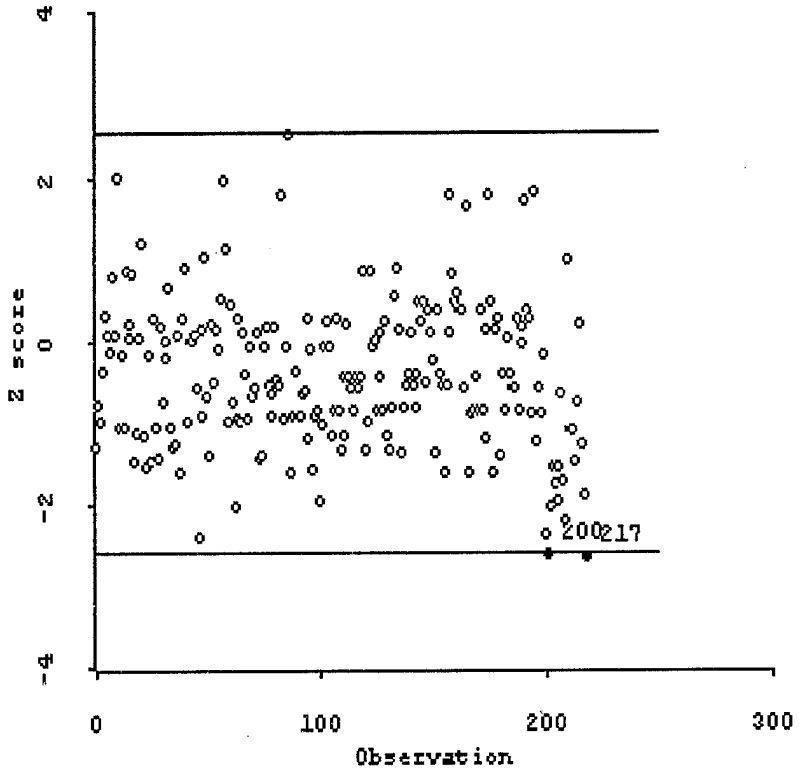


Figure 7.7: Self-starting Shewhart chart for μ. Training set has been redefined to 200 observations by sampling with replacement from the original 80 base case points. The chart signals immediately when the data from the historical model is charted. Note that the points are labeled from 0 upward, and the point plotted as observation zero is observation 3 compared with observations 1 and 2 from the data stream. The signal at point 200 corresponds to the 3rd point from the historical data.

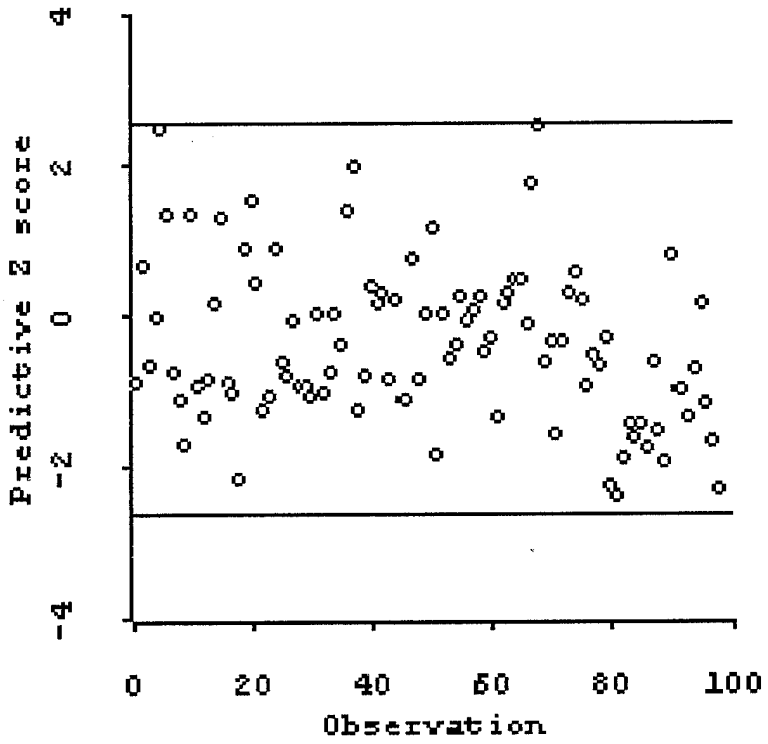


Figure 7.8: Predictive chart for the base case followed by the historical data. While the chart does not signal, evidence indicates a downward trend after observation 80.

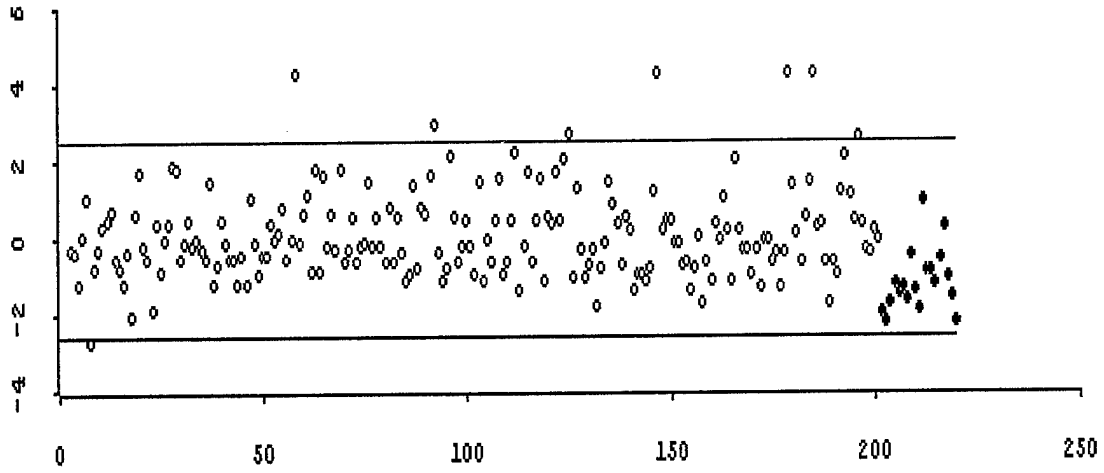


Figure 7.9: A predictive Shewhart chart with bootstrapping to increase the training set to 200 points from 80. Note the first two observations are not plotted (since we need at least two points to predict the next observation) and points that are plotted are numbered with the first point being 0, as is the *Xlisp-Stat* convention. The bold points at the right of the chart are the observations from the historical model, different from the base case. There are several interesting aspects to this chart, including the unexpectedly high number of out-of-control points. Those out-of-control points are also highlighted.

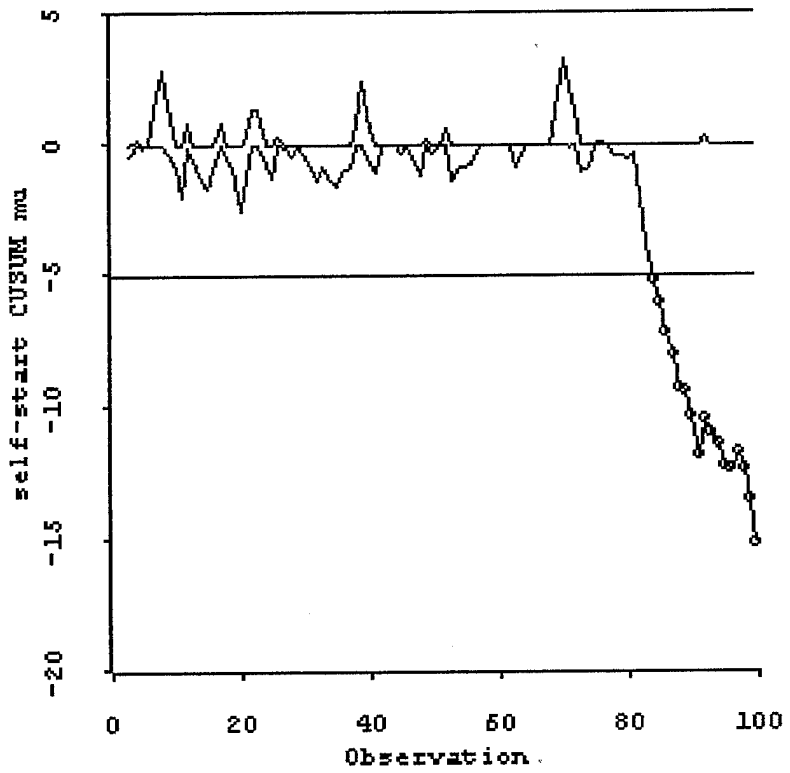


Figure 7.10: Self-starting CUSUM of base case and historical case. $ARL = 250$. Note the chart signals at observation 84, four observations after going out of control.

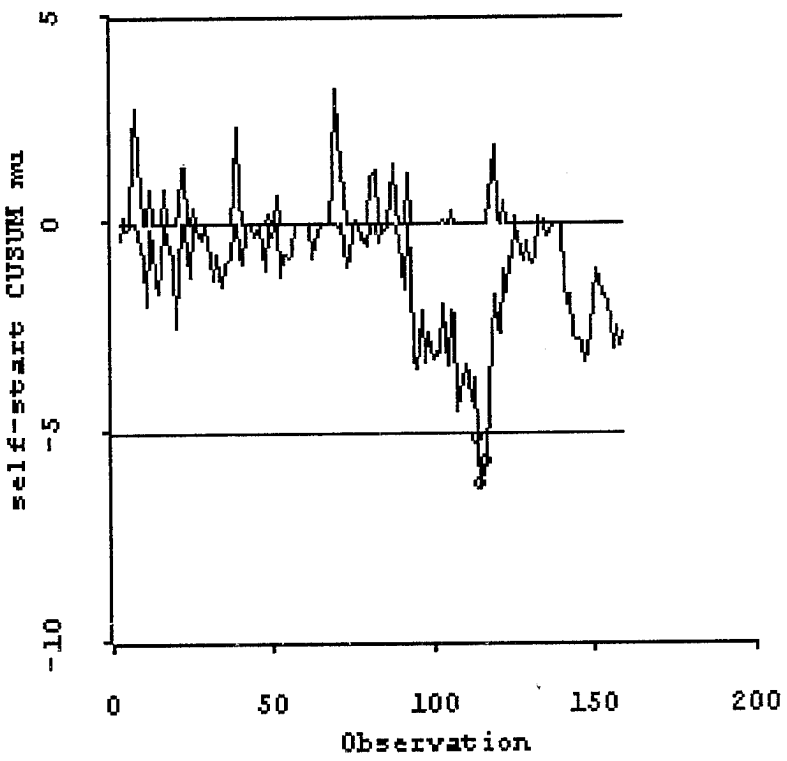


Figure 7.11: Self-starting CUSUM of base case and first model change. $ARL = 250$. Note the chart signals at observation 114.

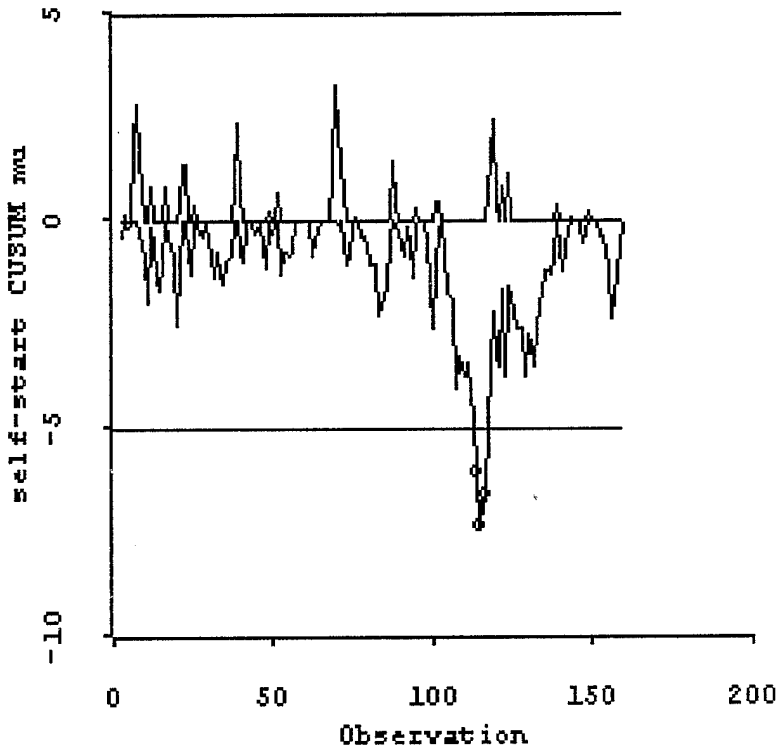


Figure 7.12: Self-starting CUSUM of base case and second model change. $ARL = 250$. Note this chart also signals at observation 114.

7.5 Conclusions

We have seen that the methods of this thesis detect changes in the output of combat simulation models. This is significant for three reasons.

First, we have modeled the output of these combat models as inverse-gaussian variates. This is a novel modeling strategy, and offers promise for fresh insight into the behavior of these combat simulations. It allows a new approach to comparing the output of different versions of the computer code based on statistical methodology, as the sampling theory of inverse gaussian variates is well known. In particular, it offers a way to determine, before simulation begins, an appropriate number of simulation runs, based on power considerations. This would be a vast improvement over the current approach of running 30 simulations due to an appeal to the central limit theorem.

Second, we have applied control chart methodology to this new modeling strategy. This holds promise for quality control of large computer programs of any sort which are subject to continual revision.

Last, we have shown how the tools of this thesis can be used in a practical setting. Our control chart methods, based on the inverse gaussian distribution, apply in this context where other methods based on other distributions would have been inappropriate.

Chapter 8

Application to automobile manufacturing: General Motors

This chapter examines another application of the inverse gaussian distribution as a model for task completion time. The example comes from an article in *Applied Statistics*, by A. F. Desmond and G. R. Chapman [1993]. We again demonstrate how the tools developed in this thesis can be useful to both the industrial statistician and quality controller.

8.1 Background

In their paper, Desmond and Chapman examined the time to complete a task by crews of workers at the General Motors plan in Oshawa, Ontario. The crews performed repetitive tasks on an assembly line. The facility in which they worked electronically monitored "all aspects of the operation of the plant."[Desmond and Chapman, 1993]. Desmond and Chapman looked at three processes, of which we consider one. The thrust of Desmond and Chapman's paper was modeling these task completion times with mixtures of inverse gaussian distributions. The third process "exhibits no mixing whatsoever." This is the process we will study in this chapter.

We examine the station where the parts required to install a radio were assembled. Desmond and Chapman[1993] describe this "radio kitting station" as the place where " a single worker performs the simple task of reading a docket indicating what parts are required, taking the parts from the appropriate bins, and placing them into

a tray."

For reasons very close to the ones addressed in our introductory chapter, Desmond and Chapman chose to model the task completion time using inverse gaussian random variables, instead of using Weibull, gamma, lognormal or other skewed non-negative distributions.

Those authors also caveat their data with the warning that the reported times were subject to variation from actual time, due to operator discretion in the operation of the mechanism which signaled task completion, and other sources. General Motors pre-processed the data prior to releasing it to Desmond and Chapman. The completion times were sorted; accordingly, no time-series analysis is directly applicable. Additionally, Desmond and Chapman screened for what they considered outliers (they termed them "bogus readings") and removed them.

What was reported in the article were the MLEs from 1955 observations of a process claimed to be well modeled by an inverse gaussian distribution. The authors offered to make the data itself available for examination, but had not forwarded it as of the date this was written. Accordingly, we accept their claim, and reserve for future work any additional goodness of fit testing.

The task completion times are modeled as inverse gaussian random variables, with $\hat{\mu} = 42.6257$ and $\hat{\lambda} = 66.282$. We note the large reported confidence interval for these estimates ($\mu \in (41.0678, 44.1501)$). The units of time were not specified in the paper, but we assume (from the context) that they are in seconds. Figure 8.1 shows the density of this distribution. We note the very heavy tail to the right. The median for this model is 32.5051, and the mode is 18.1071.

We note that conventional quality control charts based on the normal distribution are inappropriate here. Because of the heavy tail to the right, charts for centrality will never behave as expected, signaling too frequently. This situation argues for inverse gaussian charts, and illustrates that they are not of only theoretical

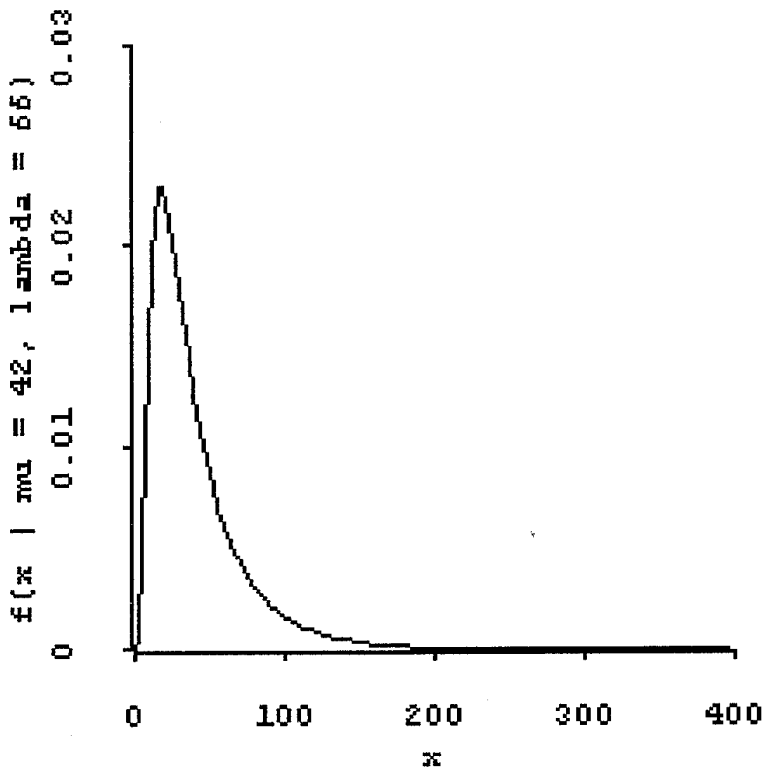


Figure 8.1: Density for an $IG(42.6257, 66.282)$ distribution.

interest.

What features are of interest to those monitoring this process? As this is a station in an assembly line, steady flow is desirable. Increases in the mean processing time or a decrease in λ, or both, will decrease the service rate at this station and can cause slowdown in the overall assembly process. On the other hand, decreases in μ and increases in λ allow management to identify circumstances which reflect improved service rates and decreased variation in the process.

We noted in earlier chapters that the HPD test was more effective at detecting increases in μ than the (corrected) UMPU test of Edgeman. Here is an instance where that appears to be advantageous.

8.2 Results

In the last chapter, we used the self-starting and predictive schemes because we had time-series data, and startup conditions. We did not use those means that assumed we had accurate process knowledge, namely the corrected Edgeman Shewhart scheme, the HPD schemes, and the standard CUSUM. For the General Motors example of this chapter, in contrast, we do not have time series data available. However, we have extensive process history (1955 observations) from which to estimate our MLEs. Accordingly, we will not use the self-starting or predictive schemes in this context. Rather, we will use the corrected Edgeman scheme, the standard CUSUM, and the HPD charts.

Between this chapter and the previous chapter, we will have illustrated each of the tools developed in this thesis.

We set the in-control ARL equal to 100 for each of the calculations that follow. We will use samples of size 1, unless otherwise indicated, to facilitate comparison between the CUSUM and Shewhart schemes.

Table 8.1: Table of ARLs for corrected Edgeman Shewhart control scheme for samples of size 1 from an $IG(\mu, 66.282)$ distribution.

ARLs out of control	
μ	ARL
20	28.8038
25	46.8314
30	66.1206
35	84.6923
40	97.8671
42.6257	100
45	98.2058
50	84.8686
60	50.5315
70	30.4222
80	20.4493

8.2.1 Shewhart Chart for the mean

We find, using Equations 2.8 and 2.9, that our control limits for process centrality are given by $6.98446 < \bar{X} < 260.141$. In control, that gives an ARL of 100. Straightforward integration gives the results in Table 8.1.

We see that even for large shifts in μ, this scheme is relatively slow to detect the process changes.

8.2.2 Shewhart chart for λ

We mentioned in Chapter 2 that the Shewhart chart for changes in λ behaved unexpectedly. We will develop that assertion now.

For this example, we consider samples of size 5. We also consider λ known from historical data, and equal to 66.282. We know that $\lambda V \sim \chi^2_4$, and we construct our control limits as Equations 2.13 and 2.14 direct.

We find $LCL = .206987$ and $UCL = 14.8602$. Those limits give an ARL in control of 100, and out-of-control ARLs as presented in Table 8.2.

We note that the scheme is not effective for detecting upward shifts in λ.

Table 8.2: Table of ARLs for the corrected Edgeman Shewhart scheme for λ for an in-control distribution of $IG(42.6257, 66.282)$. The scheme is reasonably effective detecting decreases in λ, but is not effective detecting upward shifts.

ARLs out of control	
λ	ARL
10	1.44617
20	2.89911
30	6.57186
40	15.6796
50	36.7708
60	74.6842
66.282	100
70	110.963
80	118.352
90	105.769
100	89.6385
110	75.6325
120	64.4051
130	55.4939
140	48.3559
150	42.5583

This results from both the skewness of the distribution of the χ^2 random variable and the nature of the test. An upward shift in λ corresponds to a scale change in the distribution of V, resulting in compression of the distribution of V towards zero. The skewness of the distribution means that the such compression adds more probability on the right hand side of the control chart, following such a shift, than is lost on the left hand side. Consider a shift from $\lambda = 66.282$ to $\lambda_o = 80$, an increase by a factor of about 1.2. With $\lambda_o = 80$, $P(\lambda V < .206987) = .0071840$, while $P(\lambda V > 14.8602) = .0012730$. In other words, the probability of being signaled as out-of-control has dropped to 0.00845170 from 0.01000, or about 16%. This results in the increase of the ARL to 118.352. Figure 8.2 illustrates. Since this chart is essentially the same chart for a scale change in a normal distribution, which is given by the limits

$$LCL = \frac{\chi^2_{\alpha/2,n-1}\sigma^2}{n-1} \leq S^2 \leq \frac{\chi^2_{1-\alpha/2,n-1}\sigma^2}{n-1} = UCL$$

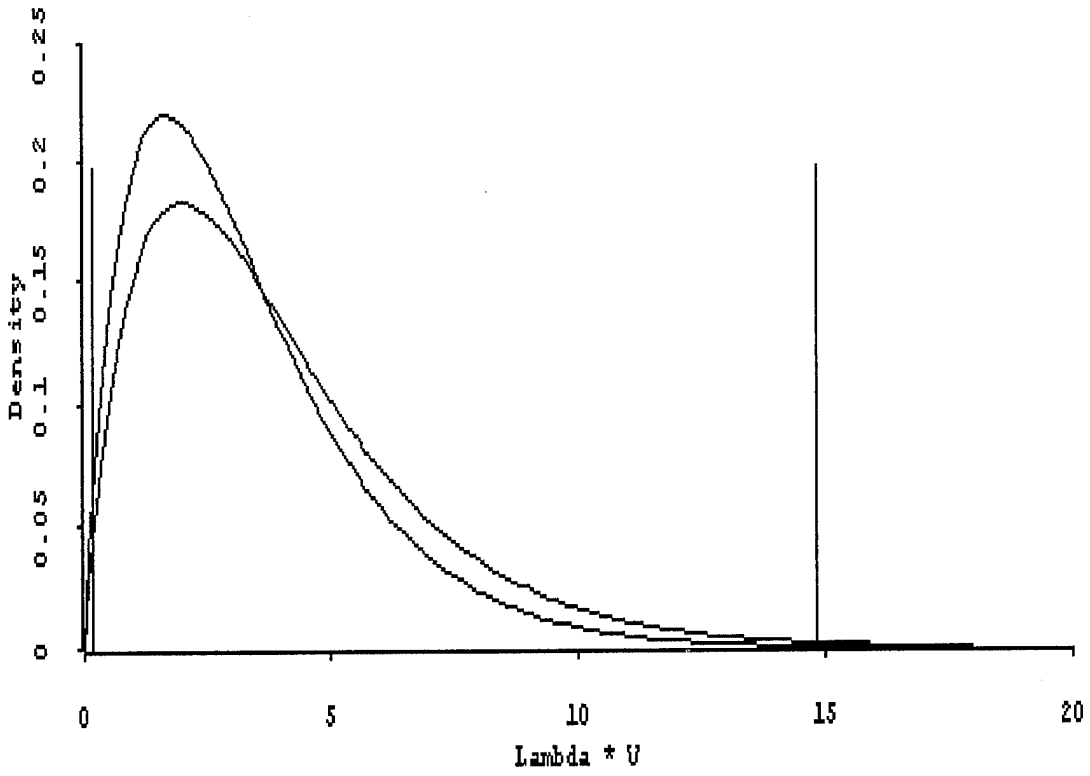


Figure 8.2: This graph shows the pdf of the sampling distribution of V when in control (the lower curve, $\lambda = 66.282$) and when λ has increased to 80. Note that the increase in λ has resulted in a compression of the density towards the origin.

it appears that using S^2 charts would suffer from the same defect.

While this scheme does not detect increases in λ well, in this context that is not necessarily a fatal defect. Recall that the variance of an IG random variable is inversely proportional to λ. Increases in λ result in decreased process variability, which is desirable. It is less critical to detect improvements quickly. On the other hand, the scheme does a reasonably good job of detecting shifts in λ which result in increased variability, and it is increased variability which will wreak havoc with a queuing system. If the scheme has to be weak in one direction, it at least is weak in the least dangerous direction.

8.2.3 HPD chart for the mean

Calculating the HPD limits for samples of size one, we obtain $LCL = 3.05051$ and $UCL = 171.599$. We compare these limits with the corrected Edgeman limits (6.98446, 260.141) and see by inspection that the HPD test will be quicker to detect upward shifts in the mean, and slower to detect downward shifts in the mean. Again, in this context, that is desirable behavior. We illustrate with Table 8.3. This example, when compared with the earlier corrected Edgeman example, highlights that the HPD test is not the uniformly most powerful test. However, since we are more interested in detecting increases in the mean than in detecting decreases, that feature is not unattractive. Yet compared to a one-sided test, it still can detect downward shifts in the mean, albeit more slowly.

8.2.4 HPD chart for λ

We find that the HPD limits for V are given by $LCL = .017469$ and $UCL = 13.2854$. Since these limits are to the left of the Edgeman limits, we expect that we will again see the HPD scheme to be more sensitive to one side than the corrected Edgeman

Table 8.3: Table of ARLs for HPD scheme for the mean of an $IG(\mu, 66.282)$ distribution with samples of size 1. We see that the test detects upward shifts much better than it detects downward shifts.

ARLs out of control	
μ	ARL
10	1064.88
20	14490
30	2115
40	152.205
42.6257	100
50	42.2147
60	20.4234
70	12.9217
80	9.46838
90	7.5729
100	6.40606
110	5.62765
120	5.07682

charts. We present Table 8.4 to illustrate.

8.2.5 CUSUM chart for the mean

We now turn our attention to CUSUM charts for the mean. We construct a table of ARLs for various shifts in the mean, and present it in Table 8.5. The ARLs presented are for a shift to the indicated out-of-control state, with the CUSUM parameters selected for maximum power to detect changes to that state. As expected, the CUSUM scheme detects these changes much more quickly than the Shewhart scheme.

8.2.6 CUSUM chart for λ

We now turn our attention to CUSUM charts for λ. We construct a table of ARLs for various shifts in λ, and present it in Table 8.6. The ARLs presented are for a shift to the indicated out-of-control state, with the CUSUM parameters selected for maximum power to detect changes to that state. Compare Table 8.6 to Table 8.2,

Table 8.4: Table of ARLs for HPD scheme for λ of an $IG(44.6257, \lambda)$ distribution with samples of size 1. We see that the test detects upward shifts much better than it detects downward shifts.

ARLs out of control	
λ	ARL
10	1.35924
20	2.46350
30	5.02293
40	10.9235
50	24.7391
60	57.581
66.282	100.0
70	136.451
80	325.887
90	769.017
100	1716.42
110	3308.50
120	4965
130	5724
140	5591
150	5115

Table 8.5: Table of ARLs for CUSUM for the mean, GM example. We use an in-control ARL of 100 for $\mu = 42.6257$ and $\lambda = 66.282$.

ARLs out of control	
μ	ARL
20	6.93
25	11.32
30	18.83
35	33.16
40	65.23
42.6257	100
45	64.42
50	34.28
60	16.54
70	10.86
80	8.23

Table 8.6: Table of ARLs for CUSUM for λ, GM example. We use an in-control ARL of 100 for $\mu = 42.6257$ and $\lambda = 66.282$. Note that this is a CUSUM of individual observations.

ARLs out of control	
λ	ARL
10	3.05
20	5.79
30	10.27
40	18.05
50	32.52
60	62.59
66.282	100
70	80.82
80	52.60
90	38.92
100	31.06
110	26.01
120	22.51
130	19.95
140	18.06
150	16.50

the table for the corrected Edgeman Shewhart charts for λ. The latter table was for samples of size five, and it performed poorly. By contrast, the CUSUM of individual observations does not display increased ARLs for out-of-control states with increased λ.

8.3 Conclusions

We have presented performance data for our various tools, applied to the General Motors case. We have seen that the CUSUM schemes for μ and λ perform much better than the HPD and Edgeman schemes. We have also seen that the HPD scheme performs better than the Edgeman scheme for detecting process changes of interest to the process manager, namely increases in μ and decreases in λ.

Finally, by presenting an assembly line application in the automobile manufacturing industry, we hope to have again shown that the methods of this thesis are

of practical as well as theoretical significance.

Chapter 9

Conclusions

This short chapter summarizes the work of this thesis, and outlines further research areas.

9.1 Summary and significance

The title of this work is "Topics in Statistical process control". In this thesis, we have explored several topics which found common application in the setting of Chapter 7 and Chapter 8.

We first reviewed the existing literature on the statistical process control of inverse gaussian variates, due to Edgeman. We extended and modified his work with Shewhart control schemes.

We introduced self-starting and predictive control charts, explored their properties, and gave examples of their use. As part of that work, we derived the predictive distribution for the next m observations of an inverse gaussian variate, given the first n observations.

We introduced HPD control regions for skewed distributions, and explored their advantages and disadvantages.

We explored bivariate control charts, and gave new diagnostic rules for classifying out of control observations. We extended this idea to bivariate control charts based on HPD regions.

We developed optimal CUSUM schemes for the parameters of an inverse gaussian distribution. As a consequence, we obtained optimal CUSUM schemes for the

scale parameter of a Gamma variate. We developed numerical routines to find the parameters of these CUSUM schemes based on both variance reduction simulation and numerical approximations to integral equations, using work of Hawkins for the latter.

We found a new application for these tools, control of software modeling military combat. We modeled the output of those simulations as inverse gaussian variates, and applied our new tools to this data. We extended this work to the idea of controlling complex software subject to constant revision.

Finally, we returned to an industrial setting for our final application, and showed that even the much examined automotive assembly line could benefit from the application of the tools of this thesis.

In all of the above work, we wrote software to implement the algorithms we developed, and used that software in our examples.

9.2 Future work

There are several exciting areas for future research opened by this thesis.

First, what is the distribution of the run length of a CUSUM scheme which goes out of control when the S^+ and S^- are not zero? Since most departures from control occur when these values are not zero, this has practical significance as well as theoretical interest. Solution allows for generalizing the idea of the Fast Initial Response CUSUM.

Second, how can the diagnostic tools for a bivariate chart be sharpened?

Third, we believe that other outputs of military combat simulations are better modeled as inverse gaussian variates, especially the time to complete a mission, the time to reach a decision, and time to move a certain distance. We wish to explore this further, with an eye to developing regression models to explain these responses.

We also believe that modeling these measures of effectiveness as inverse gaussian random variates allows rational principles of design of experiments to be applied to the conduct of these simulations, notably power and sample size calculations.

Fourth, can we elicit and apply informative prior distributions to these military models? How does that effect our power and sample size?

Fifth, can we develop bivariate CUSUM schemes with diagnostic tools similar to the bivariate Shewhart schemes?

Sixth, can we find or develop efficient numerical integration techniques for higher dimensional integrals involving indicator functions?

Last, what other skewed process could be well modeled by the inverse gaussian distribution, and benefit from the application of these tools to their control?

These questions will keep us occupied for the near future.

Appendix A

Statistical Computing

This appendix discusses the various pieces of code used to support the work done in this dissertation. We include it for two reasons. First, it allows someone to request code to check the results obtained in this dissertation. Second, it allows someone who desires to apply the results of this thesis to save the labor of devising and writing functionally equivalent algorithms.

Because of the preference at the University for *XLISP-STAT* [Tierney, 1990], much of the code is written in that language. However, to take advantage of some earlier results, other routines are written in FORTRAN using the IMSL routines [IMSL, 1989]. We understand that the computer scientist might be offended by such an eclectic approach.

This appendix is divided into functional pieces, as indicated by the headings.

A.1 Generating IG variates

This *Xlisp-Stat* function is based upon the algorithm recommended by Chhikara and Folks[1989], which in turn was based on work by Michael et al. [1976].

```
;;; generate IG's ;;; uses algorithm in Chhikara/Folks
(defun inv-gauss-rand (n mu lambda) (let* ((y (chisq-rand n 1)) (x1
(* (/ mu lambda 2) (+ (* 2 lambda) (* mu y) (- ((+ (* 4 lambda mu y) (*
mu mu y y)) .5))))) (u (uniform-rand n)))
(dotimes (i n) (setf (select x1 i) (if (< (select u i) (/ mu (+ mu
(select x1 i)))) (select x1 i) (/ (* mu mu) (select x1 i))))) x1 )) This
```

and a number of other *Xlisp-Stat* function are available from the author upon request.

A.2 CUSUM ARL FORTRAN routines with explanation

We have four classes of FORTRAN routines, available from the author. The first generates tables of ARL vs. h for either the optimal or the naive value of k. We have these routines for upward and downward shifts of μ and λ. There are 8, named ARL1.f – ARL8.f.

The second class of programs finds the value of h to obtain a given one-sided ARL for either the optimal or naive value of k. As before, we have these routines for upward and downward shifts of μ and λ. There are 8, labeled find1.f – find8.f

The third class of routines finds the cutoff values for HPD regions, using numerical integration routines. There is one routine in this class – findbik.f, which finds the value of k such that $P(f(x,y) > k) = p$.

The fourth class of routines computes the predictive p value for a series of observations, using the predictive distribution. This can be used to run a Shewhart or CUSUM chart of p-values. The routine labeled pred1.f computes the p-value for the next observation. The routine predm.f computes the p-value for the next m observations.

A.2.1 CUSARL

Many of these FORTRAN routines are based on a sub-program called CUSARL.f, by Douglas Hawkins. That routine evaluates the average run length for a one-sided CUSUM chart for an arbitrary data distribution of one parameter. We have made one slight modification, allowing for a two-parameter distribution to be evaluated (such

as the $IG(\mu, \lambda)$.

CUSARL uses a Markov chain approximation to find the ARLs, with an iterative mesh technique which uses Richardson extrapolation to incorporate the information from previous iterations into the current estimate. This algorithm is accurate and computationally efficient.

We do not reproduce the CUSARL code here; it is contained in the paper by Hawkins [1992c] and is also available by *ftp* from the StatLib archives at Carnegie Mellon University.

A.2.2 Finding ARL for shifts in μ and λ: *ARL1 - 8.f*

The routines ARL1.f —ARL4.f find tables of values for a shift in the mean of the $IG(\mu, \lambda)$ distribution. The user enters the in-control distribution, the out-of-control tuning value for the mean, and a range of h values. The programs compute the in-control and out-of-control values of the ARL for both conventional and fast initial response schemes, and write them to a file for importation into a graphics program.

These routines allow the user to see the trade-offs in using each of the schemes, and how the ARLs in-control tend to grow non-linearly with increasing h, while the out-of-control ARLs tend to grow linearly.

Each of these routines is for a point alternative, or a one-sided CUSUM scheme. ARLs for the two-sided schemes must be found by combining the ARLs as described in Equation 1.11.

ARL1.f finds the ARL for an upward mean shift using the optimal scheme. ARL2.f finds the ARL for an upward mean shift using the arithmetic mean. ARL3.f find the ARL for a downward mean shift using the optimal scheme. ARL4. f finds the ARL for a downward mean shift using the arithmetic mean.

ARL5.f through ARL8.f are the same as ARL1.4 through ARL4.f, except they

are for shifts in λ.

These routines are available from the author upon request.

A.2.3 Finding decision limits for a CUSUM scheme.

The optimal scheme in this thesis tells us what k, the reference value, should be. We still have to determine h, the reference value, to meet a desired ARL for the given k.

We used a modified Newton-Raphson method with a difference quotient instead of the exact derivative to solve for h. We used the CUSARL code to give us the ARL of an arbitrary h as $f(h) = ARL_h$, and then we found the root of the equation $f(h) - ARL = 0$ for our desired ARL.

These programs are named FINDH1.f through FINDH8.f, and correspond to the cases for the ARL code.

These, too, are available from the author.

A.3 Variance reduction techniques for IG ARLs with code

We needed a check on our programming for the FORTRAN routines. We developed simulation routines to find ARLs for our schemes. We used a clever variance reduction scheme developed by Jun and Choi [1991], and applied it to the inverse gaussian distribution. The technique uses total hazard as a control variate. The routine is available from the author as *IG_ARL.LSP*.

A.4 Data

The data files used in this thesis are available from the author.

Appendix B

BIBLIOGRAPHY

B.1 Works Cited

[1] Banerjee, Asit K. and G. K. Bhattacharyya (1979) Bayesian Results for the Inverse Gaussian Distribution with an Application. *Technometrics* Vol. 21(2). pp. 247-251.

[2] Barnard, G. A. (1959) Control Charts and Stochastic Processes. *Journal of the Royal Statistical Society, Series B.* Vol. 21. No. 2. pp. 239-257.

[3] Bather, J. A. (1963) Control Charts and Minimization of Costs. *Journal of the Royal Statistical Society, (B).* Vol. 25.

[4] Bickel, Peter J. and Kjell A. Doksum. (1977) *Mathematical Statistics.* Englewood Cliffs, NJ: Prentice Hall.

[5] Bissell, A. F. (1969) CUSUM techniques for quality control (with discussion). *Applied Statistics.* Vol. 18. PP. 1-30.

[6] Champ, C. W. and W. H. Woodall. (1987) Exact results for Shewhart control charts with supplementary runs rules. *Technometrics.* Vol. 29. pp. 393-399.

[7] Champ, C. W. and W. H. Woodall. (1990) A program to evaluate the run length distribution of a Shewhart control chart with supplementary runs rules. *Journal of Quality Technology.* Vol. 22. pp. 68-73.

[8] Chhikara, R. S. (1975) Optimum tests for the comparison of two inverse Gaussian distribution means. *Australian Journal of Statistics.* Vol. 17. pp. 77-83.

[9] Chhikara, R. S. and J. L. Folks. (1976) Optimum test procedures for the mean of first passage time in Brownian motion with positive drift (inverse Gaussian distribution). *Technometrics.* Vol. 19. pp. 189-193.

[10] Chhikara, R. S. and J. L. Folks. (1977) The Inverse Gaussian Distribution as a Lifetime Model. *Technometrics* Vol. 19. No. 4. pp. 461-468.

[11] Chhikara, R. S. and J. L. Folks. (1989) *The Inverse Gaussian Distribution.* New York: Marcel Dekker.

[12] Chhikara, Raj S. and Irwin Guttman. (1982) Prediction Limits for the Inverse Gaussian Distribution. *Technometrics* Vol. 24. No. 4. pp. 319-324.

[13] Desmond, A. F. and G. R. Chapman. (1993) Modeling Task Completion Data with Inverse Gaussian Mixtures. *Applied Statistics.* Vol. 42. No. 4. pp. 603-613.

[14] Dupuy, Trevor N. (1987) Can We Rely Upon Computer Combat Simulations? *Armed Forces Journal International.* August. pp. 58-63

[15] Edgeman, Rick L.(1989) Inverse Gaussian Control Charts. *Australian Journal of Statistics.* Vol. 31. No. 1. pp. 78-84.

[16] Edgeman, Rick L., Robert C. Scott, and Robert J. Pavur. (1988) A modified Kolmogorov-Smirnov Test for the Inverse Gaussian Density with Unknown Parameters. *Communications in Statistics - Simulations.* Vol. 17. No. 4. pp. 1203-1212.

[17] Gan, F. F. (1992a) Exact Run Length Distributions for One-sided Exponential CUSUM schemes. *Statistica Sinica.* Vol. 2. pp. 297-312.

[18] Geisser, Seymour (1993) *Predictive Inference: An Introduction* New York: Chapman & Hall.

[19] Girshick, M. A. and H. Rubin. (1952) A Bayes' Approach to a Quality Control Model. *Annals of Mathematical Statistics.* Vol. 23.

[20] Goel, A.L. (1982) Cumulative Sum Control Charts. In *Encyclopedia of Statistical Science* Edited S. Kotz and N. L. Johnson. New York: John Wiley and Sons. pp. 233-241.

[21] Hawkins, Douglas M. (1987) Self-starting cusums for location and scale. *The Statistician* Vol. 36. pp. 299-315.

[22] Hawkins, Douglas M. (1992b) *Statistical Quality Improvement* Unpublished lecture notes, Statistics 5911. University of Minnesota.

[23] Hawkins, Douglas M. (1992c) Evaluation of Average Run Lengths of Cumulative Sum Charts for an Arbitrary Data Distribution. *Communications in Statistics, Part B – Simulation and Computation* Vol. 21. pp. 1001-1020.

[24] Helmbold, Robert L. (1990) *Rates of Advance in Historical Land Combat Operations.* Bethesda, Maryland: US Army Concepts Analysis Agency.

[25] IMSL, Inc. (1989) *MATH/LIBRARY: FORTRAN Subroutines for Mathematical Applications.* Edition 1.1. Houston: IMSL, Inc.

[26] Hughes, Wayne P., editor. (1984) *Military Modeling.* Alexandria, VA: Military Operations Research Society.

[27] IMSL, Inc. (1989) *STAT/LIBRARY: FORTRAN Subroutines for Statistical Applications.* Edition 1.1. Houston: IMSL, Inc.

[28] Johnson, N. L. (1961) A Simple Theoretical Approach to Cumulative Sum Control Charts. *Journal of the American Statistical Association.* Vol. 56. pp. 835–840.

[29] Johnson, N. L. and F. C. Leone. (1962a) Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part I. *Industrial Quality Control.* Vol. 18, No. 12, pp. 15-21.

[30] Johnson, N. L. and F. C. Leone. (1962b) Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part II. *Industrial Quality Control.* Vol. 19. No. 1. pp. 29-36.

[31] Johnson, N. L. and F. C. Leone. (1962c) Cumulative Sum Control Charts: Mathematical principles applied to their construction and use. Part III. *Industrial Quality Control.* Vol. 19. No. 2. pp. 22-28.

[32] Johnson, N. L. (1966) Cumulative Sum Control Charts and the Weibull Distribution. *Technometrics.* Vol. 8. No. 3. pp. 481-491.

[33] Jun, Chi-Hyuck and Moon Soo Choi. (1993) Simulating the Average Run Length for CUSUM Schemes using Variance Reduction Techniques. *Communications in Statistics B: Simulation.* Vol. 22. No. 3. pp. 877-887.

[34] Lanchester, F. W. (1956) Mathematics in Warfare. In *The World of Mathematics.* Vol. 4. Edited J. R. Newman. New York: Simon and Schuster. pp. 2138 -2157.

[35] Lucas, James M. and Ronald B. Crosier. (1982) Fast Initial Response for CUSUM Quality Control Schemes: Give your CUSUM a Head Start. *Technometrics.* Vol. 24. No. 3. pp. 199-205.

[36] Michael, J.R., Schucany, W.R., and Haas, R. W. (1976) Generating random variables using transformation with multiple roots. *American Statistician* Vol. 30. pp. 88-90.

[37] Montgomery, Douglas C. (1985) *Introduction to Statistical Process Control.* 1st Ed. New York: Wiley.

[38] Montgomery, Douglas C. (1991) *Introduction to Statistical Process Control.* 2nd Ed. New York: Wiley.

[39] Moustakides, George V. (1986) Optimal Stopping Times for Detecting Changes in Distributions. *Annals of Statistics.* Vol. 14. No. 4. pp. 1379-1387.

[40] Page, E. S.(1954) Continuous Inspection Schemes. *Biometrics* Vol. 41. pp. 1-9.

[41] Regula, Gary A. (1976) Optimal Cumulative Sum Procedures to Detect a Change in Distribution for the Gamma Family. Ph.D. Dissertation, Case Western Reserve University.

[42] Ritov, Y.(1990) Decision Theoretic Optimality of the CUSUM Procedure. *Annals of Statistics* Vol. 18. No. 3. pp. 1464-1469.

[43] Roberts, S. W. (1959) Control Charts based on Geometric Moving Averages. *Technometrics.* Vol. 1. pp. 239-250.

[44] Ross, S. M. (1971) Quality Control under Markovian Deterioration. *Management Science.* Vol. 17.

[45] Savage, I. R. (1962) Surveillance Problems. *Naval Research Logistics Quarterly.* Vol. 9.

[46] Schroedinger, E. (1915) Zur Theorie der fall- und steigversuche an teilchen mit Brownscher bewegung. *Phys. Ze.* Vol. 16. pp. 289–295.

[47] Shewhart, Walter A. (1931) *Economic Control of Quality of Manufactured Product.* New York: Van Nostrand.

[48] Taylor, H. M. (1965) Markovian Sequential Replacement Processes. *Annals of Mathematical Statistics.* Vol. 36.

[49] Taylor, H. M. (1965) Statistical Control of a Gaussian Process. *Technometrics.* Vol. 9.

[50] Taylor, James G. (1981) *Force-on-Force Attrition Modeling.* Arlington, VA: Operations Research Society of America.

[51] Taylor, James G. (1983) *Lanchester Models of Warfare.* Volumes I and II. Arlington, VA: Operations Research Society of America.

[52] Tierney, Luke. (1990) *LISP-STAT: An Object Oriented Environment for Statistical Computing and Dynamic Graphics.* New York: Wiley.

[53] Tweedie, M. C. K. (1957a) Statistical properties of inverse Gaussian distributions I. *Annals of Mathematical Statistics.* Vol. 28. pp. 362-377.

[54] Tweedie, M. C. K. (1957b) Statistical properties of inverse Gaussian distributions II. *Annals of Mathematical Statistics.* Vol. 28. pp. 696-705.

[55] Van Dobben de Bruyn, C. S. (1968) *Cumulative Sum Tests: Theory and Practice.* London: Griffin.

[56] Varian, Hal R. (1975) A Bayesian Approach to Real Estate Assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage* Eds. Stephen Feinberg and Arnold Zellner. Amsterdam: North-Holland. pp. 195-208.

[57] Ventisel, Ve. S. (1964) *Introduction to Operations Research* Moscow: Soviet Radio Publishing House.

[58] Wald, Abraham. (1944) On Cumulative Sums of Random Variables. *Annals of Mathematical Statistics.* Vol. 15.

[59] Wald, Abraham (1945) Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics.* Vol. 16.

[60] Wald, Abraham. (1947) *Sequential Analysis* New York: Wiley.

[61] Wald, Abraham, and Jacob Wolfowitz. (1948) Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics.* Vol. 19. pp. 326-339.

[62] Western Electric (1956) *Statistical Quality Control Handbook.* Indianapolis: Western Electric Corporation.

[63] Woodall, William H. (1983) The distribution of run-length of one-sided CUSUM procedures for continuous random variables. *Technometrics* Vol. 25. pp. 295-301.

[64] Woodall, William H. (1986) The Design of CUSUM Quality Control Charts. *Journal of Quality Technology* Vol. 18. No. 2. pp. 99-102.

[65] Zellner, Arnold. (1986) Bayesian Estimation and Prediction Using Asymmetric Loss Functions. *Journal of the American Statistical Association.* Vol. 81.

No. 394. pp. 446-451.

B.2 Other references

[1] Alwan, Layth C. (1986) CUSUM Quality Control – Multivariate Approach. *Communications in Statistics B.* Vol. 15. No. 12. pp. 3531-3543.

[2] Athreya, K.B. (1986) Another Conjugate Family for the Normal Distribution. *Statistics and Probability Letters* Vol. 4. pp. 61-64.

[3] Box, G. E. P. and G. M. Jenkins. (1976) *Time Series Analysis: Forecasting and Control* 2nd ed. San Francisco: Holden-Day.

[4] Crosier, Ronald B. (1988) Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* Vol. 30. No. 3. pp. 291-303.

[5] Doganaksoy, Necip; Frederick W. Faltin; and William T. Tucker. (1991) Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment. *Communications in Statistics B: Theory and Methods.* Vol. 20. No. 9. pp. 2774-2790.

[6] Ferguson, Thomas S. (1967) *Mathematical Statistics: A Decision Theoretic Approach.* New York, Academic Press.

[7] Folks, J. L and R. S. Chhikara (1978) The Inverse Gaussian Distribution and its Statistical Application – A Review. *Journal of the Royal Statistical Society B.* Vol. 40. No. 3. pp. 263-289.

[8] Fries, Arthur. (1986) Optimal Design for an Inverse Gaussian Regression Model. *Statistics and Probability Letters.* Vol. 4. pp. 291-294.

[9] Gan, F. F. (1992b) CUSUM control charts under linear drift. *The Statistician* Vol. 41. pp. 71-84.

[10] Hawkins, Douglas M. (1992a) A Fast Accurate Approximation for Average Run Lengths of CUSUM Control Charts. *Journal of Quality Technology* Vol. 24. No. 1. pp. 37-43.

[11] Lindgren, Bernard. (1976) *Statistical Theory.* 3rd ed. New York: MacMillan.

[12] Nyhoff, Larry and Sanford Leestma. (1992) *FORTRAN 77 for Engineers and Scientists.* 3rd Ed. New York: Macmillan.

[13] Pignatiello, Joseph J. Jr. and George C. Runger. (1990)Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology.* Vol. 22 (3). pp. 173-186.

[14] Tiao, G. C. (1972) Asymptotic behavior of temporal aggregates of time series. *Biometrika.* Vol. 59. pp. 525-531.

[15] Yashchin, Emmanuel. (1993) Statistical Control Schemes: Methods, Applications, and Generalizations. *International Statistical Review* Vol. 61. No. 1. pp. 41-66.