5-2024

Information Based Approach for Detecting Change Points in Inverse Gaussian Model with Applications

Alexis Anne Wallace

Information Based Approach for Detecting Change Points in Inverse Gaussian Model with Applications

---

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

Alexis Wallace

May 2024

Information Based Approach for Detecting Change Points in Inverse Gaussian Model with Applications

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

by

Alexis Wallace

May 2024

Approved by:

Suthakaran Ratnasingam, Committee Chair

Hani Aldirawi, Committee Member

Dalton Marsh, Committee Member

Dr. Madeleine Jetter, Chair, Department of Mathematics

Dr. Corey Dunn, Graduate Coordinator

Abstract

Change point analysis is a method used to estimate the time point at which a change in the mean or variance of data occurs. It is widely used as changes appear in various datasets such as the stock market, temperature, and quality control, allowing statisticians to take appropriate measures to mitigate financial losses, operational disruptions, or other adverse impacts. In this thesis, we develop a change point detection procedure in the Inverse Gaussian (IG) model using the Modified Information Criterion (MIC). The IG distribution, originating as the distribution of the first passage time of Brownian motion with positive drift, offers flexibility and effectively models a wide range of data shapes. Moreover, it handles outliers and skewness better than some other distributions. Extensive simulation studies are conducted to illustrate the performance of our proposed method compared to existing methods across various settings, in terms of type I error, power, and confidence set. The results indicate that our MIC-based approach is comparable to the Schwarz Information Criterion method. Further, the proposed method has an advantage, especially when the change occurs at the very beginning or at the very end of the dataset. Finally, we present two real-world data applications to demonstrate the advantage of our proposed method.

Table of Contents

Abstract **iii**

List of Tables **vi**

List of Figures **vii**

1 The Inverse Gaussian Distribution **1**
1.1 Literature Review 1
1.2 The Inverse Gaussian Distribution Model 1
1.3 Properties of the Inverse Gaussian Distribution 4
1.3.1 Additivity 4
1.3.2 Exponential Family 5
1.3.3 Completeness 6
1.3.4 Reproducibility 6
1.4 Moments and Moment Generating Function 7
1.5 The Hazard Function 7
1.6 Convergence in Inverse Gaussian to Normal 8
1.7 Generating Samples from the Inverse Gaussian Distribution 9
1.8 Maximum Likelihood Estimates for the Inverse Gaussian Distribution . . 10
1.9 Comparing the Inverse Gaussian Distribution Densities as model 10

2 Change Point Analysis Using the Modified Information Criterion **13**
2.1 Literature Review 13
2.2 The Change Point Problem 14
2.3 Binary Segmentation Procedure 15
2.4 Modified Information Criterion 16
2.5 MIC-based Detection Procedure for Inverse Gaussian distribution 18
2.6 Confidence Set Calculations: Profile log-likelihood and Deviance Function 20

3 Simulation Study **22**
3.1 Simulation Study Settings 22
3.2 Type I Error Simulation for MIC and SIC 23
3.3 Power Simulation for MIC and SIC 25
3.4 Coverage Probability Simulation for MIC and SIC 31

4 Real Data Application **36**
4.1 Compressive Strength and Strain of Maize Seeds 36
4.2 Measurement of Dispersion of Particulate Matter 2.5 in Thailand 39

5 Conclusion **42**

Bibliography **44**

List of Tables

3.1 Type I Error for MIC and SIC Across Different Sample Sizes with Nominal Level $\alpha = 0.05$ 24
3.2 Power Comparison for MIC and SIC with Change in Mean with Pre-change Model $X_L \sim IG(1,1)$ 26
3.3 Power Comparison for MIC and SIC with Change in Variance with Pre-change Model $X_L \sim IG(1,1)$ 28
3.4 Power Comparison for MIC and SIC with Change in both Mean and Variance with Pre-change Model $X_L \sim IG(1,1)$ 30
3.5 Coverage Probabilities for MIC and SIC with Change in Mean for Various Sample Sizes and Different Change Point Locations 33
3.6 Coverage Probabilities for MIC and SIC with Change in Variance for Various Sample Sizes and Different Change Point Locations 34
3.7 Coverage Probabilities for MIC and SIC with Change in Mean and Variance for Various Sample Sizes and Different Change Point Locations 35

List of Figures

1.1 Probability Density Function 3
1.2 Cumulative Distribution Function 4

3.1 Type I Comparison for MIC and SIC 24
3.2 Power Comparison for MIC and SIC with Change in Mean with Pre-change Model $X_L \sim IG(1, 1)$ 27
3.3 Power Comparison for MIC and SIC with Change in Variance with Pre-change Model $X_L \sim IG(1, 1)$ 29
3.4 Power Comparison for MIC and SIC with Change in both Mean and Variance with Pre-change Model $X_L \sim IG(1, 1)$ 31

4.1 Change point detection for the Compressive Strength of Maize Seeds Using MIC 38
4.2 Change point detection for the Compressive Strength of Maize Seeds Using SIC 39
4.3 Change point detection for the measurement of Dispersion PM 2.5 Using MIC 41

Chapter 1

The Inverse Gaussian Distribution

1.1 Literature Review

The Inverse Gaussian distribution is closely connected to the Gaussian distribution, as implied by its name. The Inverse Gaussian (IG) distribution originates as the distribution of the first passage time of Brownian motion with positive drift [SS99]. However, Halphen is given credit with the first formulation of what is known as the generalized Inverse Gaussian distribution. The IG distribution name is derived from the key fact that its cumulant generating function is the inverse of the Gaussian distribution. The fundamental properties of the IG distribution were examined by and later pioneered by Tweedie, M. C. K. [Twe57]. Chhikara and Folks [CF77] explored the IG distribution as a lifetime model by looking at the model's application for studying reliability aspects given a high occurrence of early failures.

1.2 The Inverse Gaussian Distribution Model

Originally derived by Tweedie M. C. K. [Twe57], taking only positive values for the random variable, the probability density function for the Inverse Gaussian (IG) distribution can be represented as

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}}\ \exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right], \quad \text{for} \quad x > 0, \tag{1.1}$$

where μ (> 0) and λ (> 0) are the mean and the shape parameters, respectively. These parameters μ and λ are estimated by their maximum likelihood estimators which is discussed further in Section 1.7.

The IG random variable X with probability density function (1.1) is written in shorthand as $X \sim IG(\mu, \lambda)$. The expected value of the random variable X is μ ($E[X] = \mu$). The IG distribution is an exponential distribution used to model positively skewed, non-negative data which many real-world data sets exhibit this means the IG model provides a better fit for data that deviates from normality but exhibits some similarities. Ultimately, the IG model is very versatile to model data that exhibits heavy skewness and tails.

Figure 1.1 below, displays the graph of the probability density function (1.1) for the IG distribution with various values for both μ and λ. It can be seen that the IG distribution exhibits a bell-shaped curve, similar to the Gaussian distribution. However, unlike the Gaussian distribution, the tails of the IG distribution decay more rapidly. The mean is given by μ, the shape parameter is given by λ, and the variance is determined by $\frac{\mu^3}{\lambda}$. That is, when that ratio becomes small, the IG model becomes highly skewed. Alternatively, when the ratio gets larger, the model tends to be more symmetric. Thus, as the ratio tends to infinity, the IG model becomes asymptotically normal (with mean μ and variance $\frac{\mu^3}{\lambda}$) which as mentioned previously makes the IG distribution a good model for normally distributed data.

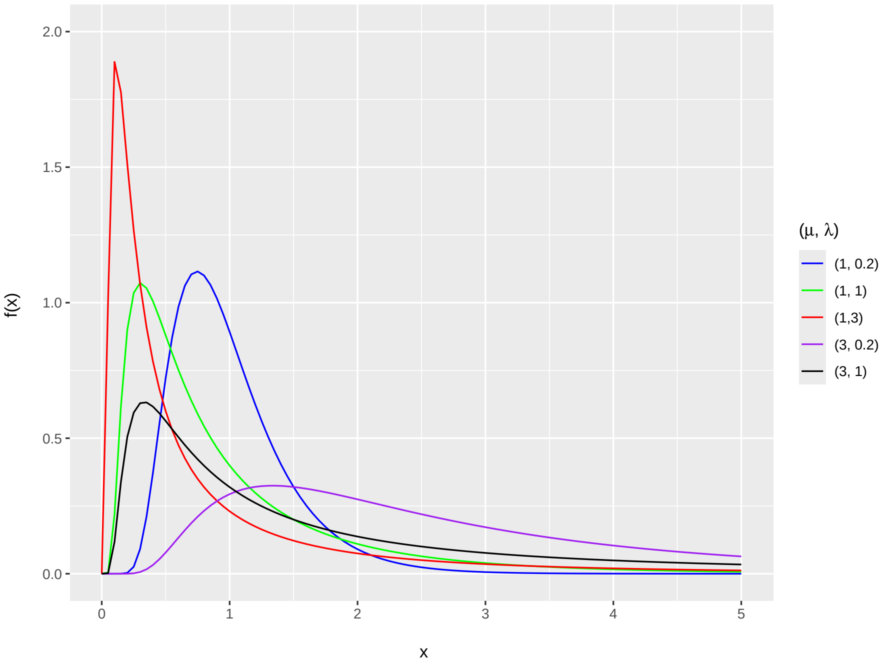

Figure 1.1: Probability Density Function

Chhikara and Folks [CF77] found that the cumulative distribution function for the IG distribution is given by the following,

$$F(x; \mu, \lambda) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(-\frac{x}{\mu}\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(1 + \frac{x}{\mu}\right)\right), \quad \text{for} \quad x > 0, \tag{1.2}$$

where Φ denotes the cumulative distribution function of the standard normal distribution, which is given by $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} \exp(-\frac{\xi^2}{2}) d\xi$. Figure 1.2 displays the graph of the cumulative distribution function for the IG distribution with various values for both μ and λ.

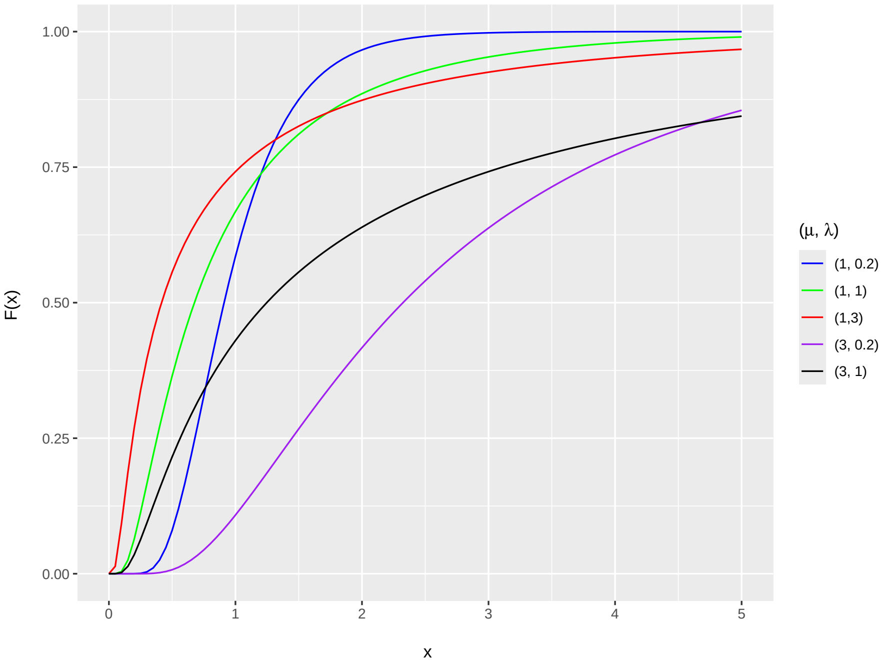

Figure 1.2: Cumulative Distribution Function

1.3 Properties of the Inverse Gaussian Distribution

As mentioned by Chhikara and Folks [CF77], the Inverse Gaussian (IG) distribution shares many similarities with the Gaussian distribution (the bell curve). However, unlike the Gaussian distribution which typically describes the level of a Brownian motion at a specific time, the IG distribution characterizes the time it takes for a Brownian motion with a positive drift to reach a predetermined positive level. So, it's the "inverse" in the sense that it deals with time rather than level. The IG distribution possesses several unique properties, including additivity, being part of the exponential family, completeness, and reproducibility which are all discussed further below.

1.3.1 Additivity

Following Sato and Inoue [SI93], linear combinations (the sum of variables) follow the Inverse Gaussian (IG) distribution. That is, letting $X \sim IG(\mu_i, \lambda_i)$ with $i = 1, 2, \cdots$, if $\frac{\lambda_i}{\mu_i^2 c_i} = \xi$ (constant) then it holds that the linear sum $\sum c_i X_i \sim IG(\sum c_i \mu_i,\ \xi(\sum c_i \mu_i)^2)$.

If we consider a linear combination of IG random variables from the probability density function (1.1) such that $Y = \sum_i c_i X_i$. And, given that $\frac{\lambda_i}{\mu_i^2 c_i} = \xi$ then $\lambda_i = \xi \mu_i^2 c_i$. Then we can calculate both the mean and variance of Y to show the IG distribution follows the additive property. To calculate the mean of Y, we have the following,

$$E(Y) = E(\sum_i c_i X_i) = \sum_i c_i E(X_i) = \sum_i c_i \mu_i.$$

Now to calculate the variance of Y, we will have the following,

$$Var(Y) = \sum_i c_i^2 Var(X_i) + 2 \sum_{i<j} c_i c_j Cov(X_i, X_j).$$

The covariance term will go away as the random variables are independent meaning their covariance will be 0. Therefore, the $Var(Y) = \sum_i c_i^2 Var(X_i)$. Substituting $Var(X_i) = \frac{\mu_i^3}{\lambda_i}$ and $\lambda_i = \xi \mu_i^2 c_i$ in $Var(Y)$, gives us $\frac{1}{\xi} \sum_i c_i \mu_i$. Therefore, Y follows an IG distribution with $\mu_Y = \sum_i c_i \mu_i$ and $\lambda_Y = \frac{1}{\xi} \sum_i c_i \mu_i$ since both the mean and variance of Y satisfy the pasteurization of the IG distribution. We can conclude that Y follows the additive property.

1.3.2 Exponential Family

The Inverse Gaussian (IG) distribution has a deep connection within the exponential family. The IG distribution belongs to the exponential family of distributions of the form from Equation 1.1. However, as noted by Sato and Inoue [SI93], Equation 1.1 can be rewritten as,

$$f(x; \theta_1, \theta_2) = \sqrt{\frac{\theta_1}{2\pi}} exp(\sqrt{\theta_1 \theta_2}) x^{(-3/2)} \exp\left[-\frac{1}{2}(\theta_1 x^{-1} + \theta_2 x) \right], \tag{1.3}$$

where x and x^{-1} are the sufficient statistics and $-\frac{1}{2}\theta_1$ and $-\frac{1}{2}\theta_2$ are the natural parameters. This implies that $T = (X_{-1}, X)$ where $X_{-1} = \frac{1}{X}$ which form a sufficient statistic for the IG distribution. Using the statistics, we get the probability function,

$$f(X_{-1}, X; \theta_1, \theta_2) = h(X_{-1}, X) \exp(-\frac{1}{2}(\theta_1 X_{-1} + \theta_2 X)),$$

which is a probability density function in the exponential family. Therefore, the sufficient statistics T with the parameter vector $\theta = -\frac{1}{2}(\theta_1, \theta_2)$ for the IG distribution follows the exponential distribution.

1.3.3 Completeness

As mentioned by Sato and Inoue [SI93], the family of Inverse Gaussian (IG) distributions is complete meaning that since the expectation of any function of the data under the IG distribution is zero for both of the parameters then the function itself is 0 everywhere.

Definition 1.1. *A statistic T is called complete if for every parameter θ, the expected value of any function of T is zero under the distribution, then the probability that the function of T equals zero for all θ is one. [Pat16]*

Say $X \sim IG(\mu, \lambda)$. By the completeness property, $\int_0^{\infty} g(x) * h(x) * f(x; \mu, \lambda) dx = 0$ for all values of μ and λ, then it must be the case that $h(x) = 0$ almost everywhere. Thus, the IG distribution can represent any possible continuous distribution. If there was a non-zero function $h(x)$ that existed such that the integral was zero for all μ and λ, then it would imply that the distribution could be more precisely modeled by adjusting both parameters. However, that is not the case and the IG distribution is a complete representation within its family of distributions.

1.3.4 Reproducibility

Noted by Sato and Inoue [SI93], if $X \sim IG(\mu, \lambda)$, then the characteristic function $C_X(t)$ is derived as the following,

$$C_X(t) = E[e^{itX}] = exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 - \frac{2it\mu^2}{\lambda}\right)^{1/2}\right]\right\}. \tag{1.4}$$

Positive and negative moments exist for the Inverse Gaussian (IG) characteristic function which can be found in detail in the following subsection. In Equation 1.4, cX is given by $\exp\left\{\frac{\lambda}{\mu}\left[1 - \left(1 - \frac{2it\mu^2 c}{\lambda}\right)^{1/2}\right]\right\}$, for an arbitrary constant $c > 0$ and assuming $X \sim IG(\mu, \lambda)$. Pulling from the expression above, with the parameters $c\mu$ and $c\lambda$, cX follows the IG distribution. Thus, for position transformations, the IG distribution is not closed, that is $(X + c)$. Implying that, for scale transformations (cX), the IG distribution is closed.

1.4 Moments and Moment Generating Function

As noted in Section 1.2, both positive and negative moments exist for the characteristic function (1.4). These moments can be obtained by calculating the moment-generating function. The moment-generating function can be found by integrating the Inverse Gaussian (IG) characteristic function (1.4). Thus, the moment-generating function is the following,

$$\begin{aligned} M_X(t) &= \int_0^\infty \exp(tx)\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right]dx && (1.5)\\ &= \exp\left\{\frac{\lambda}{\mu}\left[1-\sqrt{1-\frac{2\mu^2 t}{\lambda}}\right]\right\}. && (1.6)\end{aligned}$$

Then we can calculate the r^{th} moment by using the following equation,

$$E[X^r] = \mu^r \sum_{s=0}^{r-1} \frac{(r-1+s)!}{s!(r-1-s)!}\left(\frac{2\lambda}{\mu}\right)^{-s}.$$

This expression gives a way to compute moments of various orders for the IG distribution. Around 0, the first two moments are μ and $\mu^2 + \frac{\mu^3}{\lambda}$ while the moments around the center μ_i are $\mu_2 = \frac{\mu^3}{\lambda}$ and $\mu_3 = 3\frac{\mu^5}{\lambda^2}$. Looking at higher-order moments can give more insight into the shape and variability of the distribution giving a better understanding of its statistical properties. The relation between positive and negative moments is the following,

$$E[X^{-r}] = \frac{E[X^{r+1}]}{\mu^{2r+1}}, \tag{1.7}$$

where X^{-1} is considered as the negative moment. The skewness (β_1) and kurtosis (tailedness, β_2) of X are $\sqrt{\beta_1} = 3\sqrt{\mu/\lambda}$ and $\beta_2 = 15\mu/\lambda + 3$, respectively. The relationship between β_1 and β_1 is modeled as the following,

$$\beta_2 = \frac{5}{3}\beta_1^2 + 3, \tag{1.8}$$

and is used to determine whether data is derived from the IG distribution. Moments and moment-generating functions of other distributions, such as normal distribution or others of the exponential family can give good insight into model selection.

1.5 The Hazard Function

Common in reliability analysis, as mentioned by Chen and Gupta [CG12], a common change point problem deals with estimating the change point in a hazard function. Hazard rate functions ultimately determine the chance of an event occurring at any given point in time. Thus, the Inverse Gaussian (IG) distribution is particularly useful in estimating change points within hazard functions. Chhikara and Folks [CF77] determined the hazard rate for the IG distribution is defined as,

$$h(x,\theta) = \frac{(\frac{\lambda}{2\pi x^3})^{1/2}\exp(-\frac{\lambda(x-\mu)^2}{2\mu^2 x})}{\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right)+\exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}\right)+1\right)},$$

with a given point in time, x, and parameters μ and λ. This IG hazard rate function exhibits an unimodal pattern, rising from 0 to its maximum value before gradually declining asymptotically to a constant. For this reason, the IG distribution is commonly used in reliability and survival analysis.

1.6 Convergence in Inverse Gaussian to Normal

Sato and Inoue [SI93] discuss how under certain conditions the Inverse Gaussian (IG) distribution converges to the normal distribution. Let $X \sim IG(\mu,\lambda)$ and $N \sim N(m,\sigma)$. For $Y = \frac{\sqrt{\lambda}(X-\mu)}{\mu\sqrt{X}}$ and $Z = \frac{N-m}{\sigma}$, the following properties hold.

- The standard variable Y, defined above, for the IG distribution follows the standard seminormal distribution which indicates that Y closely resembles the standard normal distribution.

- Y^2 and Z^2, both defined above, exhibit a χ^2 (continuous) distribution with one degree of freedom showing their shared statistical behavior in terms of variability.

- The arithmetic mean $\bar{X}$ of IG random variables X_i follows $IG \sim (\mu, n\lambda)$. $\bar{X}$ serves as both the maximum likelihood estimator and the minimum variance unbiased estimation of μ. Similarly, the sample mean $\bar{Z}$ of normal random variables Z_i follows $N \sim (m, \frac{\sigma^2}{n})$, and $\bar{Z}$ serves as the maximum likelihood estimator and the minimum variance unbiased estimation of m. This similarity shows a significant convergence between the IG distribution and the normal distribution implying that methodologies developed for the normal distribution can be applied to approximate and analyze data from the IG distribution.

- For n independent and identically distributed IG random variables X_i, $i = 1, 2, \cdots n$, the sum $\sum_{i=1}^{n} X_i$ and $W = \sum_{i=1}^{n} X_i^{-1} - \bar{X}^{-1}$ are independent, with W following the χ^2 distribution with $(n-1)$ degrees of freedom. Similarly, for n normal random variables Z_i, the sample variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ follows the χ^2 distribution with $(n-1)$ degrees of freedom.

- The maximum likelihood estimation (MLE) and minimum variance unbiased estimation (MVUE) have similar patterns for both distributions, the IG and normal distributions. Specifically, for the IG distribution, the MLE of $\frac{1}{\lambda}$ is $\frac{1}{n}W$, while the MVUE is $\frac{1}{n-1}W$. Conversely, for the normal distribution, the MLE of the parameter σ^2 is $\frac{1}{n}S^2$ and the MVUE is $\frac{1}{n-1}S^2$.

1.7 Generating Samples from the Inverse Gaussian Distribution

Using the method by Sato and Inoue [SI93], we can generate a sample sequence following the Inverse Gaussian (IG) distribution. We first must generate a random number from the χ_1^2 distribution with one degree of freedom. This distribution is characterized by a probability density function, $f(y) = \frac{1}{2\sqrt{y}}e^{-y/2}$ for $y > 0$. One of the properties from Section 1.5 explains that $Y^2 = \frac{\lambda(X-\mu)^2}{\mu^2 X}$, where X follows the IG distribution, then Y^2 follows the χ^2 distribution. Next, we solve the equation $\frac{\lambda(X-\mu)^2}{\mu^2 X} = Y^2$ for X. This equation yields a quadratic equation with two roots, denoted as X_1 and X_2. These roots can be calculated using the formulas,

$$X_1 = \frac{\mu}{2\lambda}\left[2\lambda + \mu Y^2 - \sqrt{4\lambda\mu Y^2 + \mu^2 Y^4}\right] \text{ and,}$$
$$X_2 = \frac{\mu^2}{X_1}.$$

Once we have X_1 and X_2, we select one of them as the generated random number. This selection is based on probabilities. Specifically, X_1 is chosen with probability $\frac{\mu}{\mu+X_1}$, and X_2 is chosen with probability $\frac{X_1}{\mu+X_1}$. This process is iterated as needed to generate a sequence of random numbers following the IG distribution. Each iteration produces a new random number from the distribution, ensuring that the generated sequence forms a sample sequence from the IG Distribution.

1.8 Maximum Likelihood Estimates for the Inverse Gaussian Distribution

First explored by Tweedie [Twe57], however following Chhikara and Folks [CF77], taking a random sample $X_1, X_2, \cdots, X_n$, the log-likelihood corresponding to $IG(\mu, \lambda)$ is

$$\ell(\mu, \lambda) = \frac{n}{2} \log\left(\frac{\lambda}{2\pi}\right) - \frac{3}{2}\sum_{i=1}^{n} \log(x_i) - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{x_i}.$$

To find the maximum likelihood estimates (MLEs) of the parameters μ and λ, we compute the partial derivatives $\frac{\partial \ell}{\partial \mu}$ and $\frac{\partial \ell}{\partial \lambda}$ and set them equal to zero. The ML estimators for μ and λ are then given by,

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

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

The conditional distribution is the same for all $\bar{X}$ as the conditional moment generating function of $n\lambda\hat{\lambda}^{-1}$ given $\bar{X}$ is $(1-2t)^{-(n-1)/2}$. Thus, the MLEs, $\hat{\mu}$ and $\hat{\lambda}$, are statistically independent from one another. The MLE of the variance, $\frac{\mu^3}{\lambda}$, is given by,

$$\hat{\sigma}^2 = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\bar{X}^3}{X_i} - n\bar{X}^2\right].$$

The MLE of the variance for the Inverse Gaussian distribution closely resembles the MLE of the variance for the normal distribution, $\frac{1}{n}\sum_{j=1}^{n}(x_j - \hat{\mu})^2$, as mentioned in Section 1.6.

1.9 Comparing the Inverse Gaussian Distribution Densities as model

The Inverse Gaussian (IG) distribution can be compared to other distributions as shown in the following lemmas.

Lemma 1.2. *If $X \sim IG(\mu, \lambda)$, then for any arbitrary positive value a, $aX \sim IG(a\mu, a\lambda)$.*

Proof. Let X follow an IG distribution, that is $X \sim IG(\mu, \lambda)$. We aim to show that for any arbitrary positive constant a, the random variable aX follows the IG distribution, $IG(a\mu, a\lambda)$.
Since X follows an IG distribution, its mean and variance are μ and $\frac{\mu^3}{\lambda}$, respectively.
Now, consider aX. By the properties of expectation and variance for scaled random variables, we have $E[aX] = aE[X] = a\mu$ and $Var[aX] = a^2Var[X] = a^2\frac{\mu^3}{\lambda} = \frac{(a\mu)^3}{a\lambda}$.
Therefore, aX follows the IG distribution with parameters $a\mu$ and $a\lambda$, as was to be shown. □

Lemma 1.3. *If* $X_i \sim IG(\mu, \lambda)$, *then* $\sum_{i=1}^{n} X_i \sim IG(n\mu, n^2\lambda)$.

Proof. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables following an IG distribution with parameters μ and λ. We aim to show that $\sum_{i=1}^{n} X_i$ follows an IG distribution with parameters $n\mu$ and $n^2\lambda$.
Using the moment generating function derived in Section 1.4, we have

$$M_X(t) = E(e^{tx}) = e^{\frac{\lambda}{\mu}\left(1-\sqrt{\frac{2\mu^2 t}{\lambda}}\right)}.$$

Then,

$$\begin{aligned}
M_{X_1+X_2+\cdots+X_n}(t) &= E[e^{t(X_1+X_2+\cdots+X_n)}] \\
&= E[e^{tX_1} * e^{tX_2} * \cdots * e^{tX_n}] \\
&= E[e^{tX_1}] * E[e^{tX_2}] * \cdots * E[e^{tX_n}] \\
&= [E[e^{tX_1}]]^n \\
&= [M_X(t)]^n \\
&= \left[e^{\frac{\lambda}{\mu}\left(1-\sqrt{\frac{2\mu^2 t}{\lambda}}\right)}\right]^n \\
&= e^{n\left(\frac{\lambda}{\mu}\left(1-\sqrt{\frac{2\mu^2 t}{\lambda}}\right)\right)} \\
&= e^{\frac{n\lambda}{\mu}\left(1-\sqrt{\frac{2\mu^2 t}{n\lambda}}\right)}
\end{aligned}$$

This moment-generating function matches the form of the moment-generating function (1.5) of an IG distribution with parameters $n\mu$ and $n^2\lambda$. Therefore, $\sum_{i=1}^{n} X_i$ follows the IG distribution with parameters $n\mu$ and $n^2\lambda$, as desired. □

Lemma 1.4. *If $X_i \sim IG(\mu, \lambda)$ for $i = 1, 2, \cdots n$, then $\bar{X} \sim IG(\mu, n\lambda)$.*

Proof. Let $X_1, X_2, \cdots, X_n$ be independent random variables following an IG distribution with parameters μ and λ. We aim to show that $\bar{X}$, the sample mean of these variables, follows an IG distribution with parameters μ and $n\lambda$.
Since $\bar{X}$ follows an IG distribution then we have,

$$E[\bar{X}] = E\left[\frac{\sum X_i}{\mu}\right] = \frac{1}{n}\sum E[X] = \frac{1}{n} \times n\mu = \mu,$$

and

$$Var[\bar{X}] = Var\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n^2}\sum Var[X_i] = \frac{1}{n^2} \times n\left(\frac{\mu^3}{\lambda}\right) = \frac{\mu^3}{n\lambda}.$$

Therefore, $\bar{X}$ follows the IG distribution with parameters μ and $n\lambda$. □

Lemma 1.5. *If $X_i \sim IG(\mu_i, 2\mu_i^2)$ then $\sum_{i=1}^n X_i \sim IG(\sum_{i=1}^n \mu_i, 2\sum_{i=1}^n \mu_i^2)$.*

Proof. Let X_i follow an IG distribution with parameters μ_i and $2\mu_i^2$. That is, $X_1 \sim IG(\mu_1, 2\mu_1^2)$, $X_2 \sim IG(\mu_2, 2\mu_2^2), \cdots, X_n \sim IG(\mu_n, 2\mu_n^2)$. It suffices to show that the summation of all the following random variables follows an IG distribution with parameters $\sum_{i=1}^n \mu_i$ and $2\sum_{i=1}^n \mu_i^2$.
Since X_i follows an IG distribution then we have,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \mu_i,$$

and

$$Var\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n Var[X_i] = 2\sum_{i=1}^n \mu_i^2.$$

Therefore, $\sum_{i=1}^n X_i$ follows an IG distribution with parameters $\sum_{i=1}^n \mu_i$ and $2\sum_{i=1}^n \mu_i^2$, as was to be shown. □

Chapter 2

Change Point Analysis Using the Modified Information Criterion

2.1 Literature Review

Change point analysis is very relevant as changes appear in all sorts of data sets such as quality control, the stock market, and temperature. It is important for statisticians to recognize these changes as it allows them to take the appropriate measures to best avoid and minimize losses. Change point analysis plays an important role in identifying points in time when the probability distribution of stochastic processes or time series changes. When a change point exists, it is not advisable to perform a statistical analysis without taking into account of the existence of that change point because it could lead to misleading results. Change point analysis attempts to identify the number of change point(s) and the corresponding location(s). Change point analysis has been extensively explored since 1954 by Page, E.S. [Pag54, Pag55]. Chernoff and Zacks (1964) [CZ64] estimated the current mean of a normal distribution which was subjected to changes in time by observing an object's successive positions traveling on a path. A Bayesian approach was taken here in order to better test whether a change in mean occurred in the estimation problem at hand. Sen and Srivastava (1975) [SS75a, SS75b] tested whether the means of each variable in a sequence of independent random variables can be taken to be the same, against alternatives that a shift might have occurred after some point. They also took a Bayesian approach enabling them to derive exact and asymptotic distribution functions for testing of their test statistic of a single change in the mean of a sequence of normal random variables. Srivastava and Worsley (1986) [SW86] studied a sequence of independent multivariate normal vectors with equal but possible unknown variance matrices to have equal mean vectors. They tested after an unknown change point in the sequence to see whether the mean vectors had changed. Later, Tweedie M. C. K. [Vos81] proposed the binary segmentation procedure which detects multiple structural changes recursively, saving a great deal of computing time. The general description for change point analysis can be found in the book, *Parametric Statistical Change Point Analysis* [CG12]. A multitude of other literature relating to the change point(s) problem exists, however, none have studied the change point detection procedure in the Inverse Gaussian (IG) model using the Modified Information Criterion.

2.2 The Change Point Problem

Following Chen, J. and Gupta, A. K. [CG12], the typical change point problem lets $x_1, x_2, \ldots, x_n$ be a sequence of independent random vectors with probability distribution functions $F_1, F_2, \ldots, F_n$, respectively. Then, generally, the change point problem tests the null hypothesis,

$$H_0 : F_1 = F_2 = \cdots = F_n \tag{2.1}$$

versus the alternative:

$$H_1 : F_1 \cdots = F_{k_1} \neq F_{k_1+1} = \cdots F_{k_2} \neq F_{k_2+1} = \cdots F_{k_q} \neq F_{k_q+1} = \cdots = F_n, \tag{2.2}$$

where $1 < k_1 < k_2 < \cdots k_q < n, q$ is the unknown number of change points and $k_1, k_2, \cdots, k_q$ are the respective unknown positions that have to be estimated.

In general, the change point problem involves hypothesis testing and parameter estimation. More specifically, we need to test the null hypothesis of no change point versus the alternative hypothesis of having at least one change. Further, we need to estimate the corresponding location of the change point if there are any. One of the most popular methods of detecting change points is the use of model selection criteria. The Schwarz Information Criterion (SIC) [Sch78] is one of the popular criteria for model selection. Zhang, N. and Siegmund, D. [ZS07] noted that the conventional SIC could detect change points more effectively when changes take place in the middle of the data. However, Chen, J. and Gupta, A. K. [CGP06] pointed out that the conventional SIC method did not consider the complexity of the model which may cause the redundancy of the parameter space, especially a change occurring near the beginning or the end of data. To tackle this issue, Chen, J. and Gupta, A. K. [CGP06] proposed the Modified Information Criterion (MIC) by adjusting the penalty term in SIC so that it reflects the contributions of change-point locations to model complexity. This approach assigns a larger penalty term when the change point location is close to the first or the last observation in the data set. Many existing methods can be used in change point analysis. Bayesian tests, non-parametric tests, stochastic process tests, and maximum likelihood ratio tests are among the few common, widely used tests. In this thesis, the objective is to develop a new change point detection procedure for the Inverse Gaussian model and establish the corresponding asymptotic properties.

2.3 Binary Segmentation Procedure

As proposed by Vostrikova [Vos81], in the case that multiple change points exist, a common method used is the binary segmentation procedure. This method simplifies a multiple change point problem into a single change point problem. The binary segmentation procedure saves a great amount of computing time by detecting the number and location of the change points simultaneously.

Say $x_1, x_2, \cdots, x_n$ is a random sequence of n random variables where each random variable x_i has a probability density function $F_i(\theta_i)$. When testing the null hypothesis versus the alternative hypothesis, the binary segmentation procedure along with a detection method is used to look for all possible change points (if any exist). A general summary below describes the general method of the binary segmentation procedure.

Start by assuming there is at most one change point in the probability distribution functions. So, test the null hypothesis, $H_0 : F_1 = F_2 = \cdots = F_n$, versus the alternative, $H_a : F_1 = F_k \neq F_{k+1} = \cdots = F_n$ where k is the location of the first change point. If H_0 is rejected then a change does occur at the kth observation. However, if H_0 is not rejected, then we do not have enough evidence to reject the null hypothesis (say that there is a significant change point in the data). If no change point exists, then the binary segmentation procedure stops and we come to the conclusion that no change point exists.

However, if a change point does exist, then using similar hypothesis testing, we test the two subsequences that were derived before. The test is again the null hypothesis, $H_0 : F_1 = F_2 = \cdots = F_k$, versus the alternative, $H_a : F_1 = F_{k_1} \neq F_{k_1+1} = \cdots = F_k$ where k_1 is the possible location of the change point in the first subsequence. The same procedure is followed for the second subsequence. If change points are continued to be found, then the data can continue to be separated and tested until there are no longer any change points found in each subsequence.

2.4 Modified Information Criterion

The change point problem requires us to compare the null hypothesis (2.1) which assumes no change points and the alternative hypothesis (2.2) which assumes at least one change point. With the change point problem, if change point(s) do exist, it is imperative to find the location(s). There are many methods to detect whether change points exist or not.

Let $x_1, \cdots, x_n$ be a random sample drawn from the density function $f(x; \Theta)$. The Schwarz Information Criterion (SIC) proposed by Schwarz [Sch78] is given by,

$$\text{SIC} = -2\ell_n(\hat{\Theta}) + \dim(\hat{\Theta})\log(n) \tag{2.3}$$

where $\ell_n(\cdot)$ is the log-likelihood function of the random sample, $\hat{\Theta}$ is the maximum likelihood estimator (MLE) of the parameter Θ, and $\dim(\hat{\Theta_k})$ is the dimension of the parameter space. We denote the pre-change and post-change parameters as Θ_L, Θ_R, respectively. And, we denote, $\hat{\Theta}_L, \hat{\Theta}_R$ as the MLEs of the pre-change and post-change parameters, respectively. Now, the SIC under the alternative hypothesis in the context of having at least one change point can be represented by the following,

$$\text{SIC}(k) = -2\ell_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \Big\{2\dim(\hat{\Theta}_L(k)) + 1\Big\}\log(n), \tag{2.4}$$

where $1 \leq k < n$. Again, in the above, we denote the pre-change and post-change parameters as Θ_L, Θ_R, respectively. And, we denote, $\hat{\Theta}_L, \hat{\Theta}_R$ as the MLEs of the pre-change and post-change parameters, respectively. Chen et al. [CGP06], noted that in $SIC(k)$ (2.3), the change point location is not considered as a parameter, and this absence could lead to redundancy in the parameter space if the change occurs close to the data's beginning or end. The result of this issue brought about the Modified Information Criterion (MIC). MIC has a modification of the penalty term used in SIC. This modification adjusts the penalty term in SIC so that it reflects the contributions of the change-point location to model complexity. MIC assigns a larger penalty when the change point location is close to the first one or the last observation in the data set. Under the null hypothesis, h_0, with the assumption of no change points, SIC(n) and MIC(n) are the same. We can define the MIC as,

$$\text{MIC}(n) = -2\ell_n(\hat{\Theta}) + \dim(\hat{\Theta})\log(n), \tag{2.5}$$

where $\hat{\Theta}$ maximizes $\ell_n(\Theta)$. Comparing, $SIC(n)$ (2.3) and MIC(n) (2.5), it is evident that there are no differences under H_o. However, under the alternative hypothesis, H_1, the MIC is defined as,

$$\text{MIC}(k) = -2\ell_n(\hat{\Theta}(k), \hat{\Theta}_R(k), k) + \left\{ 2\dim(\hat{\Theta}_L(k) + \left(\frac{2k}{n} - 1\right)^2 \right\} \log(n), \tag{2.6}$$

where $1 \leq k < n$. As mentioned previously, MIC(k) (2.6) differs from SIC(k) (2.4) as MIC(k) considers the contribution of the change point location k to the model parameter. It is important to note that if MIC(n) $> min_{1 \leq k < n}$MIC(k), then the model with a change point is selected. The change point is estimated by,

$$\hat{k} = arg \min_{1 \leq k < n} \{\text{MIC}(k)\}. \tag{2.7}$$

The associated MIC-based test statistic is defined as,

$$S_n = \text{MIC}(n) - min_{1 \leq k < n}\text{MIC}(k) + \dim(\Theta)\log(n), \tag{2.8}$$

which assesses the statistical significance of the identified change point. This test statistic is constructed based on the MIC of the null and the alternative hypotheses. Chen et al. [CGP06] pointed out, under Wald conditions and the regularity conditions, that as $n \to \infty$,

$$S_n \to \chi^2_d, \tag{2.9}$$

in distribution under H_o, where d is the dimension of Θ.

2.5 MIC-based Detection Procedure for Inverse Gaussian distribution

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent random variables belonging to a two-parameter Inverse Gaussian (IG) distribution. The change point problem for a two-parameter IG distribution is defined as follows.

$$X_i \sim \begin{cases} IG\ (\ \mu_L, \lambda_L\ ), i = 1, \cdots, k \\ IG\ (\ \mu_R, \lambda_R\ ), i = (k+1), \cdots, n. \end{cases} \tag{2.10}$$

where the probability density function of the IG distribution is given in 1.1. We will test the null hypothesis,

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \\ \lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda, \end{aligned}$$

versus the alternative hypothesis,

$$\begin{aligned} H_1 : \underbrace{\mu_1 = \cdots = \mu_k}_{\mu_L} \neq \underbrace{\mu_{k+1} = \cdots = \mu_n}_{\mu_R} \\ \underbrace{\lambda_1 = \cdots = \lambda_k}_{\lambda_L} \neq \underbrace{\lambda_{k+1} = \cdots = \lambda_n}_{\lambda_R}, \end{aligned}$$

where $(\mu, \lambda), (\mu_L, \lambda_L)$, and (μ_R, λ_R) are unknown parameters and need to be estimated along with the parameter k which represents the unknown change point location. Under the null hypothesis, the log-likelihood function is given as,

$$\ell_n(\mu, \lambda) = \frac{n}{2}\log\lambda + \frac{n\lambda}{2\mu^2} - \frac{n\lambda\bar{x}}{\mu^2} - \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - \mu)^2}{\lambda x_i}. \tag{2.11}$$

The maximum likelihood estimators (MLEs) of μ and λ can be obtained by setting the following partial derivatives equal to zero.

$$\begin{aligned} \frac{\partial \ell_n(\mu, \lambda)}{\partial \mu} &= \frac{n\lambda}{\mu^3} - \frac{n\lambda\bar{x}}{\mu^2} + \sum_{i=1}^{n}\frac{x_i - \mu}{\lambda x_i} \\ \frac{\partial \ell_n(\mu, \lambda)}{\partial \lambda} &= \frac{n}{2\lambda} + \frac{n\mu^2}{2\lambda^2} - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - \mu)^2}{\lambda {x_i}^2} \end{aligned}$$

The MIC(n) is defined as,

$$\begin{aligned}\text{MIC}(n) &= -2\ell_n(\hat{\mu},\hat{\lambda}) + \left[\text{dim}(\hat{\mu} + (\text{dim}(\hat{\lambda})\right]\log(n) \\ &= -2\ell_n(\hat{\mu},\hat{\lambda}) + 2\log(n),\end{aligned} \tag{2.12}$$

where $\hat{\mu}$ and $\hat{\lambda}$ are the MLEs of μ and λ, respectively. Similarly, under the alternative hypothesis, the log-likelihood function is,

$$\begin{aligned}\ell_{H_1} &= l(k,\mu_L,\lambda_L,\mu_R,\lambda_R) \\ &= \sum_{i=1}^{k}\log(f(x_i,\mu_L,\lambda_L) + \sum_{i=1}^{k+1}\log(f(x_i,\mu_R,\lambda_R) \\ &= \left\{\frac{k}{2}\log\lambda_L + \frac{k\lambda_L(1-\bar{x})}{\mu_L^2} - \frac{k}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{k}\frac{(x_i-\mu_L)^2}{\lambda_L x_i}\right\} + \\ &\left\{\frac{n-k}{2}\log\lambda_R + \frac{(n-k)\lambda_R}{2\mu_R^2} - \frac{(n-k)\lambda_R\bar{x}}{\mu_R^2} - \frac{(n-k)}{2}\log(2\pi) - \frac{1}{2}\sum_{i=k+1}^{n}\frac{(x_i-\mu_R)^2}{\lambda_R x_i}\right\}\end{aligned} \tag{2.13}$$

The MLEs of the pre-change parameters, μ_L and λ_L, can be obtained by setting the following partial derivatives equal to zero.

$$\begin{aligned}\frac{\partial\ell_{H_1}}{\partial\mu_L} &= \frac{k\lambda_L}{{\mu_L}^3} - \frac{k\lambda_L\bar{x}}{{\mu_L}^3} + \sum_{i=1}^{k}\frac{x_i-\mu_L}{\lambda_L x_i} \\ \frac{\partial\ell_{H_1}}{\partial\lambda_L} &= \frac{k}{2\lambda_L} - \frac{k{\mu_L}^2}{2{\lambda_L}^2} - \frac{1}{2}\sum_{i=1}^{k}\frac{(x_i-\mu_L)^2}{\lambda_L x_i}\end{aligned}$$

The MLEs of the post-change parameters, μ_R and λ_R, can be obtained by setting the following partial derivatives equal to zero.

$$\begin{aligned}\frac{\partial\ell_{H_1}}{\partial\mu_R} &= \frac{-(n-k)\lambda_R}{{\mu_R}^3} + \frac{(n-k)\lambda_R\bar{x}}{{\mu_R}^3} + \sum_{i=k+1}^{n}\frac{x_i-\mu_R}{\lambda_R x_i} \\ \frac{\partial\ell_{H_1}}{\partial\lambda_R} &= \frac{(n-k)}{2\lambda_R} + \frac{(n-k){\mu_R}^2}{2{\lambda_R}^2} + \sum_{i=k+1}^{n}\frac{(x_i-\mu_R)^2}{(\lambda_R x_i)^2}\end{aligned}$$

Now, MIC(k) is given by,

$$\text{MIC}(k) = -2\ell(k,\hat{\mu}_L,\hat{\lambda}_L,\hat{\mu}_R,\hat{\lambda}_R) + \left\{4 + \left(\frac{2k}{n} - 1\right)\right\}\log n, \tag{2.14}$$

where $(\hat{\mu}_L, \hat{\lambda}_L)$ are MLEs of the parameters before the change and $(\hat{\mu}_R, \hat{\lambda}_R)$ are MLEs of the parameters after the change. If there is a change, the change point k is estimated by the equation of $\hat{k}$, 2.7.

2.6 Confidence Set Calculations: Profile log-likelihood and Deviance Function

Based on MIC, we can construct a confidence curve for the change point in a two-parameter Inverse Gaussian (IG) distribution. Pointed out by Cunen et al. [CHH18] through the use of confidence distributions (CD), confidence curves along with confidence sets for change points can be estimated. A CD uses a sample-dependent distribution function on a given parameter space to estimate a parameter of interest [CHH18]. The confidence curve for the IG model can be obtained by calculating the profile log-likelihood function and the deviance function.

Maximizing the log-likelihood function (2.13) over the parameters for each possible value of k where $1 \leq k < n-1$ will result in the profile log-likelihood function. Given by,

$$\begin{aligned} \ell_{prof}(k) &= \max\Big(l\ (k, \mu_L, \lambda_L, \mu_R, \lambda_R) \Big) \\ &= l\ (k, \hat{\mu}_L, \hat{\lambda}_L, \hat{\mu}_R, \hat{\lambda}_R), \end{aligned}$$

where $\hat{\mu}_L, \hat{\lambda}_L, \hat{\mu}_R$ are the estimates to the left and right sides for a given k, respectively. Then the estimated change point location $\hat{k}$ is given by $l_{prof}(\hat{k}) = \max_k(l_{prof}(k))$. After $\hat{k}$ is obtained, the deviance function is given by,

$$D(k, x) = 2\{\ell_{prof}(\hat{k}) - \ell_{prof}(k)\}. \tag{2.15}$$

In the deviance function, $x = (x_1, \cdots, x_n)$ and $x_1, \cdots, x_k$ is a sample coming from the distribution $IG(\mu_L, \lambda_L)$ and $x_{k+1}, \cdots, x_n$ coming from $IG(\mu_R, \lambda_R)$. Now, $\ell_{prof}(\hat{k}) = \max_{1 \leq k < n}(\ell_{prof}(k))$. $\hat{k}$ is obtained through 2.7.

Using k, a confidence curve can be constructed based on the deviance function. Consider the estimated distribution of $D(k, X)$ at position k,

$$\Psi_k(X) = p_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, X) < x\}, \tag{2.16}$$

with $x \in \mathbb{R}$. Wilk's theorem states that for the case of continuous parameters, $\Psi_k(X)$ is approximately the distribution function of a χ_1^2. The confidence curve can be constructed as the following,

$$cc(k, x_{obs}) = \Psi_k(D(k, x_{obs})) = p_{k,\hat{\Theta}_L,\hat{\Theta}_R}\{D(k, X) < D(k, x_{obs})\} \tag{2.17}$$

For a discrete parameter, k, Wilks theorem does not hold, and thus $\Psi(k)$ can be computed through the simulations. The likelihood of $cc(k, x_{obs}) < \alpha$, given the true change-point, is often approximated as α. Subsequently, confidence sets for k can be visualized using the plot $cc(k, x)$, referred to as a confidence curve. The function $cc(k, x)$ represents the acceptance probability for k, or one minus the p-value for testing that particular k, where the deviance-based test rejects for high values of $D(k, x_{obs})$. The computation of Ψ_k and, consequently, $cc(k, x)$ involves simulation, expressed as

$$cc(k, x_{obs}) = \frac{1}{B} \sum_{j=1}^{B} I D(k, x_j) < D(k, x_{obs}) \tag{2.18}$$

for a large number B of simulated data sets X. This process is carried out for each candidate value of k, with simulated data X_i^* from $f(x, \Theta_L)$ and $f(x, \Theta_R)$ to the left and right sides of k respectively. Further details can be found in Cunen et. al [CHH18].

In our proposed method for constructing confidence curves, we depart from the approach in [CGH18] for estimating k. Instead of determining the change location k by maximizing the profile-likelihood function over all possible values of k, we utilize the MIC in (2.7). The MIC-based statistics S_n in (2.8) are employed to verify a statistically significant change and mitigate fluctuations caused by noise. According to Wilks' theorem, the approximate confidence curve is obtained by transforming the deviance function to the χ_1^2 probability scaling:

$$cc(k) = \Gamma_1(D(k, x)) \tag{2.19}$$

where $D(k, x)$ is defined in (2.15).

Chapter 3

Simulation Study

3.1 Simulation Study Settings

In this section, we conduct Monte Carlo simulations to evaluate the performance of the change-point detection procedure for the Inverse Gaussian model. To assess the effectiveness of the proposed method, we consider three commonly used criteria for evaluating change-point detection procedures.

1. Type I error rate: Close to the nominal level.

2. Power of the test: Preferably close to 1.

3. Coverage Probability: Close to the confidence level.

We considered the following three settings. These settings are used for power and coverage probability analysis.

- Setting 1: Mean change
  - Pre-change distribution: $IG(1,1)$
  - Post-change distribution: $IG(1.25,1), IG(1.5,1), IG(1.75,1), IG(2,1)$
- Setting 2: Variance change
  - Pre-change distribution: $IG(1,1)$
  - Post-change distribution: $IG(1,1.25), IG(1,1.5), IG(1,1.75), IG(1,2)$

- Setting 3: Mean and Variance change
  - Pre-change distribution: $IG(1,1)$
  - Post-change distribution: $IG(1.25, 1.25), IG(1.5, 1.5), IG(1.75, 1.75), IG(2, 2)$

For all three settings, we considered various sample sizes (n) and various change point locations (k).

- Case 1: $n = 50$; $k = 15, 20, 25$
- Case 2: $n = 100$; $k = 25, 40, 50$
- Case 3: $n = 200$; $k = 50, 75, 100$

3.2 Type I Error Simulation for MIC and SIC

As explained by Shreffler [Shr23], a Type I error in research occurs when the null hypothesis is erroneously rejected, leading to the incorrect assertion of significant differences that, in reality, do not exist. This situation is commonly referred to as false positives, where researchers mistakenly claim differences between groups or variables. The results presented in Table 3.1 below, provide a detailed comparison of Type I error rates across different sample sizes for both Modified Information Criterion (MIC) and Schwartz Information Criterion (SIC).

As displayed, across all sample sizes, MIC consistently demonstrates lower Type I error rates compared to SIC. For instance, at $n = 50$, MIC exhibits a Type I error rate of 0.132, while SIC has a higher rate of 0.146. This trend persists across the range of sample sizes, with MIC consistently showing a more conservative approach to avoiding false positives. This pattern suggests that MIC may be a preferable choice in scenarios where controlling Type I errors is crucial. Moreover, as the sample size increases, both MIC and SIC exhibit a general trend of decreasing Type I error rates. Notably, at the largest sample size of $n = 500$, MIC achieves a remarkable Type I error rate of 0.039, whereas SIC shows a slightly higher rate of 0.055.

These results are shown graphically in Figure 3.1. It can be seen that MIC consistently performs closer to the nominal level. Overall, the results indicate that MIC is more often sensitive to the risk of Type I errors than SIC across the range of sample sizes tested. When aiming for a balance between sensitivity and controlling Type I errors, MIC emerges as a more suitable choice. These findings underscore the critical role of method selection in minimizing the risk of false positive conclusions in change point detection analyses.

Table 3.1: Type I Error for MIC and SIC Across Different Sample Sizes with Nominal Level $\alpha = 0.05$

n	MIC	SIC
50	0.132	0.146
100	0.088	0.120
150	0.069	0.094
200	0.063	0.068
300	0.052	0.067
400	0.056	0.050
500	0.039	0.055

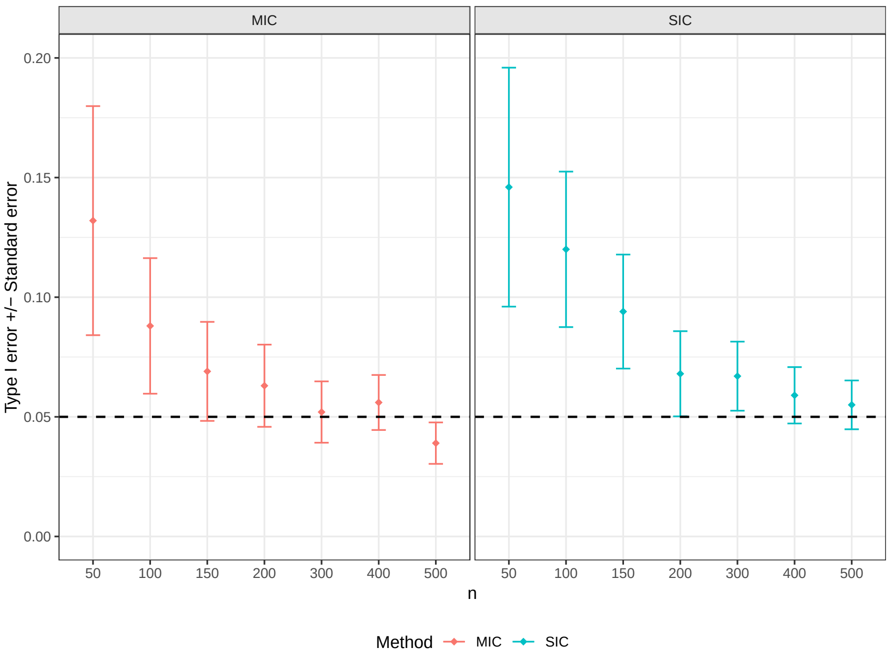

Figure 3.1: Type I Comparison for MIC and SIC

3.3 Power Simulation for MIC and SIC

Shreffler [Shr23] mentions how statistical power plays a large role in the research, particularly during the formulation and preparation stages of studies. However, it necessitates careful evaluation when interpreting outcomes. Power refers to the capacity to accurately reject a null hypothesis that is, in fact, false. Power denotes the likelihood of a study making accurate decisions or identifying an effect if it truly exists. Tables 3.2, 3.3, and 3.4 present the power comparisons between Modified Information Criterion (MIC) and Schwartz Information Criterion (SIC) with different sample sizes, parameter values, and change point locations for change in the mean, change in variance, and change in mean and variance, respectively.

In Table 3.2, the power comparisons between MIC and SIC for changes in mean are presented across different sample sizes, true change point locations (k), and parameter values. The impact of sample size on the power of detecting a change in the mean is apparent, with a noticeable increase in power as the sample size grows. For instance, in the (1.5,1) scenario, the power of both MIC and SIC improves from a sample size of 50 to 100 and continues to increase with 200 samples. The influence of the change point location is apparent as well, particularly in the (2,1) case, where both methods have higher power as the change point moves further into the later half of the dataset. The conclusions can be seen in Figure 3.2. MIC demonstrates higher power across all sample sizes and change point locations. The robustness of MIC across diverse conditions suggests its reliability for detecting changes in mean. These findings show the importance of considering the choice of the method, favoring MIC when prioritizing higher power in change point detection tasks related to shifts in the mean.

In Table 3.3, the power analysis for changes in variance across various sample sizes (n), true change point locations (k), and parameter values are presented. Again, here MIC consistently exhibits higher power compared to SIC across different sample sizes, change point locations, and parameter values. This data is presented in Figure 3.3. The results suggest that, in the context of detecting changes in variance, MIC may be a more powerful method compared to SIC. In Table 3.4, the power comparison for changes in both mean and variance is shown across varying sample sizes (n), true change point locations (k), and parameter values. The analysis provides insights into their respective abilities to detect simultaneous changes in both mean and variance. Again, MIC consistently outperforms SIC across different scenarios. This data can be seen in Figure 3.4. The results emphasize the efficacy of MIC in detecting changes in both mean and variance compared to SIC offering valuable guidance for researchers in selecting appropriate methodologies for comprehensive change point detection.

Table 3.2: Power Comparison for MIC and SIC with Change in Mean with Pre-change Model $X_L \sim IG(1,1)$

		(1.25,1)		(1.5,1)		(1.75,1)		(2,1)	
n	k	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.794	0.198	0.884	0.303	0.935	0.442	0.961	0.554
	20	0.809	0.237	0.879	0.366	0.928	0.496	0.956	0.616
	25	0.810	0.230	0.890	0.357	0.936	0.519	0.962	0.652
100	25	0.901	0.216	0.961	0.392	0.993	0.566	0.996	0.716
	40	0.895	0.207	0.960	0.440	0.987	0.664	0.995	0.801
	50	0.897	0.230	0.962	0.460	0.991	0.676	0.997	0.803
200	50	0.948	0.223	0.993	0.533	1.000	0.794	1.000	0.910
	75	0.968	0.242	0.995	0.613	1.000	0.883	1.000	0.967
	100	0.963	0.288	0.995	0.661	0.999	0.898	1.000	0.974

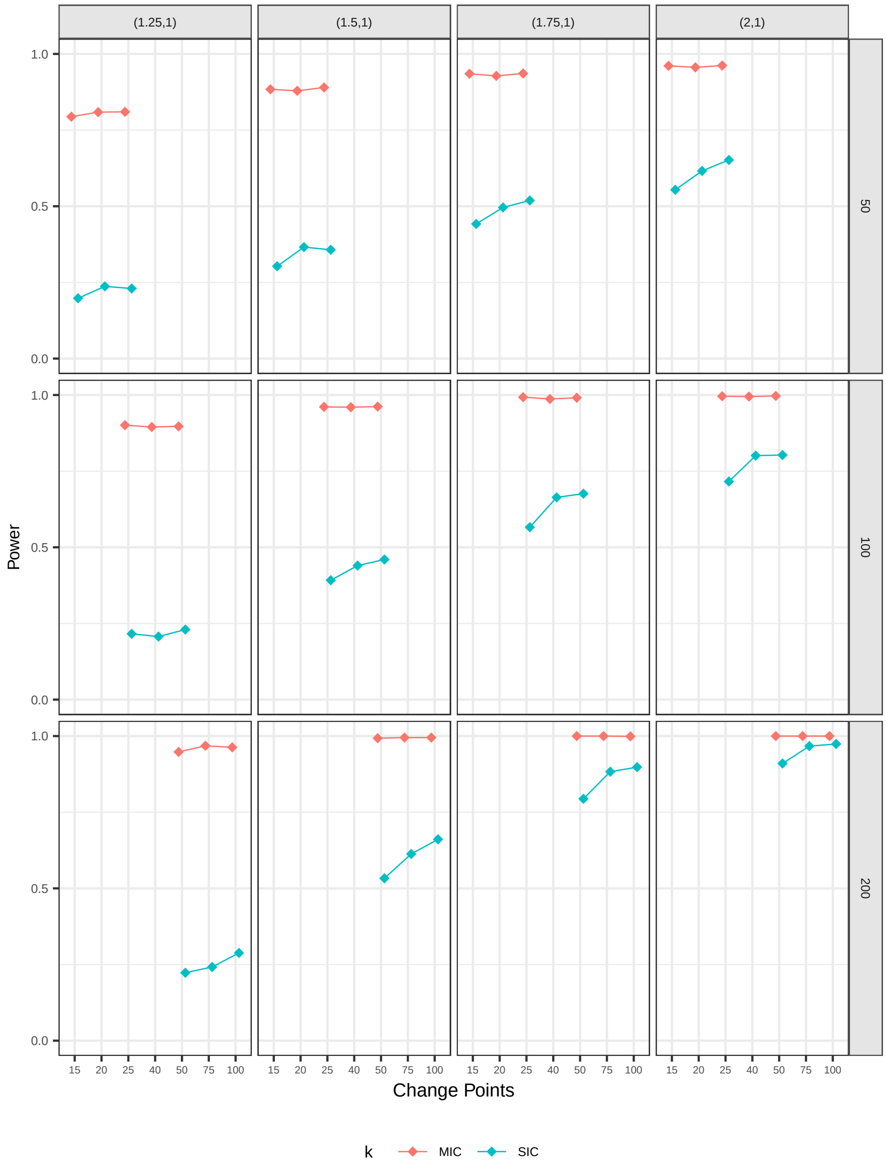

Figure 3.2: Power Comparison for MIC and SIC with Change in Mean with Pre-change Model $X_L \sim IG(1,1)$

Table 3.3: Power Comparison for MIC and SIC with Change in Variance with Pre-change Model $X_L \sim IG(1, 1)$

	(1,1.25)		(1,1.5)		(1,1.75)		(1,2)		
n	k	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.760	0.182	0.823	0.283	0.881	0.384	0.903	0.479
	20	0.793	0.214	0.846	0.296	0.884	0.407	0.926	0.519
	25	0.797	0.195	0.860	0.302	0.896	0.435	0.934	0.549
100	25	0.881	0.177	0.93	0.306	0.968	0.463	0.981	0.625
	40	0.881	0.182	0.934	0.354	0.961	0.557	0.984	0.708
	50	0.880	0.183	0.948	0.341	0.976	0.553	0.988	0.716
200	50	0.935	0.172	0.974	0.430	0.991	0.669	0.999	0.824
	75	0.943	0.193	0.983	0.471	1.000	0.744	1.000	0.903
	100	0.942	0.186	0.988	0.495	1.000	0.758	1.000	0.906

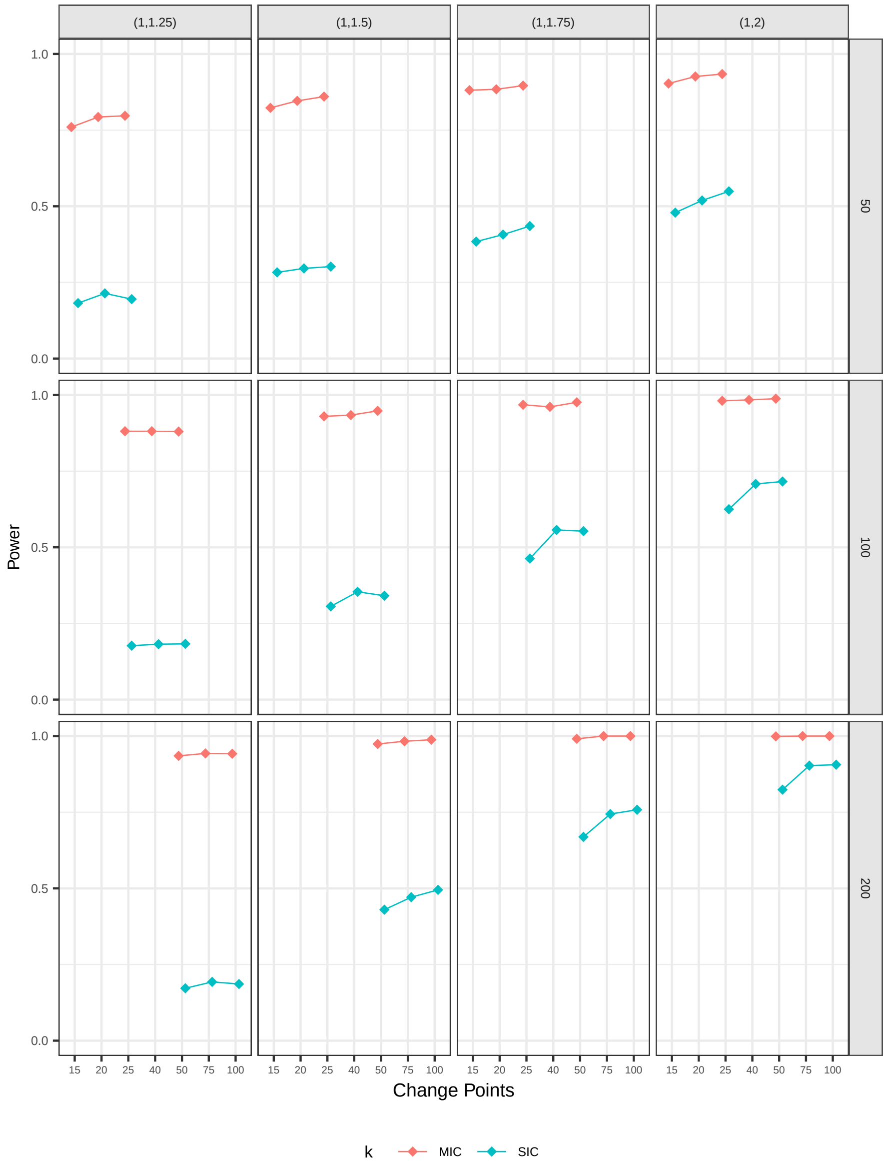

Figure 3.3: Power Comparison for MIC and SIC with Change in Variance with Pre-change Model $X_L \sim IG(1,1)$

Table 3.4: Power Comparison for MIC and SIC with Change in both Mean and Variance with Pre-change Model $X_L \sim IG(1, 1)$

		(1.25,1.25)		(1.5,1.5)		(1.75,1.75)		(2,2)	
n	k	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.844	0.266	0.934	0.491	0.974	0.719	0.990	0.860
	20	0.833	0.299	0.937	0.547	0.981	0.746	0.997	0.887
	25	0.848	0.283	0.946	0.550	0.992	0.772	0.996	0.902
100	25	0.921	0.287	0.984	0.624	0.998	0.862	1	0.953
	40	0.937	0.297	0.990	0.705	1	0.923	1	0.982
	50	0.942	0.321	0.991	0.734	1	0.933	1	0.990
200	50	0.973	0.381	0.999	0.817	1	0.974	1	0.999
	75	0.978	0.402	0.999	0.895	1	0.996	1	1
	100	0.977	0.430	1	0.905	1	0.998	1	1

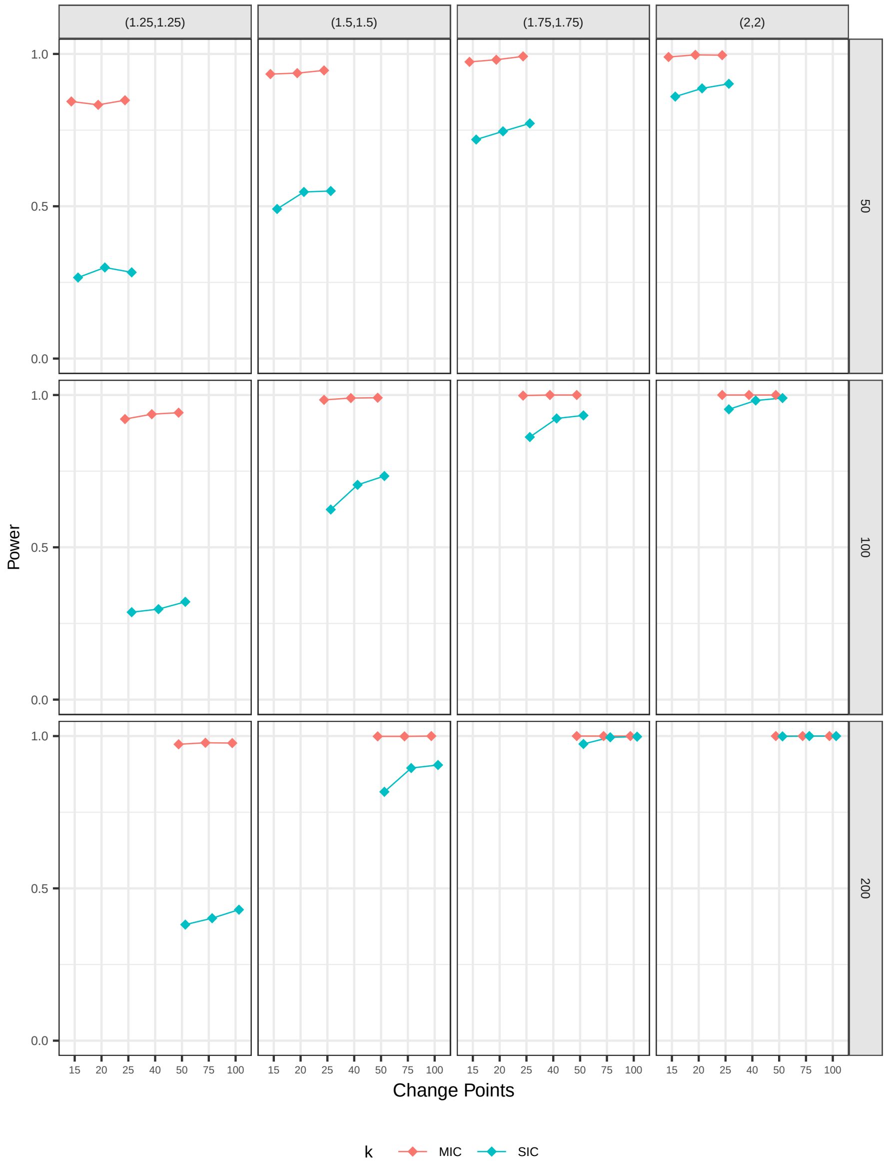

Figure 3.4: Power Comparison for MIC and SIC with Change in both Mean and Variance with Pre-change Model $X_L \sim IG(1,1)$

3.4 Coverage Probability Simulation for MIC and SIC

In this section, we conduct simulations at various values of the change point location k with different sample sizes $n = 50, 100, 200$. The pre-change distribution is always set to be $IG(1,1)$ and the post-change distribution after the change point k are generated from $IG(\mu,\lambda)$ where $\mu = \{1.25, 1.5, 1.75, 2\}$ and $\lambda = \{1.25, 1.5, 1.75, 2\}$. For the sample size $n = 50$, changes were set up to occur at $k = \{15, 20, 25\}$. For the second sample size $n = 100$, the changes were set up to occur at $k = \{25, 40, 50\}$. And for the third sample size $n = 200$, the changes were set at $k = \{50, 75, 100\}$. The changes were selected at the points to test scenarios at the beginning, middle, and end of data. Our goal is to estimate the change location $\hat{k}$ to calculate the deviance function $D(k,x)$ (2.15). We will do this using two different approaches in the following simulations. The first approach is based on the Modified Information Criterion (MIC) given in (2.4). The second approach is based on the Schwartz Information Criterion (SIC) given in (2.6). Tables 1, 2, and 3 test for change in mean, change in variance, and change in both, respectively. Testing these three changes in the parameters is important as it allows us to capture various aspects of distributional changes. It will also provide us with a better understanding of the underlying characteristics of how the data evolves over time, leading to a more comprehensive and accurate identification of the change point(s).

Table 3.5 lists out the simulation results of coverage probabilities for $n = 50, n = 100$, and $n = 200$ at two confidence levels, 0.95 and 0.99. Here we are testing change in means to identify shifts in the main tendency of the distribution. In general, MIC tends to provide higher coverage probabilities compared to SIC for the given scenarios. Both criteria show an increase in coverage probabilities as the sample size (n) increases, indicating improved performance with larger sample sizes. The impact of changing the parameters (μ, λ) is evident in the variations in coverage probabilities.

Now, table 3.6 lists out the simulation results of coverage probabilities for $n = 50, n = 100$, and $n = 200$ at two confidence levels, 0.95 and 0.99. However, here we are testing changes in variance in an attempt to identify changes in the spread of the distribution. Similar to the change in mean analysis, MIC tends to provide higher coverage probabilities compared to SIC for the given scenarios related to the change in variance. As the sample size (n) increases, both MIC and SIC show an improved coverage probability across all scenarios, meaning better performance with larger sample sizes. The impact of changing the parameters (μ, λ) is evident in the variations in coverage probabilities.

Lastly, table 3.7 lists out the simulation results of coverage probabilities for $n = 50, n = 100$, and $n = 200$ at two confidence levels, 0.95 and 0.99. However, here we are testing change in mean and variance to capture more complex shifts in the distribution as some change point events may involve simultaneous changes in both central tendency and variability. Again, MIC tends to provide higher coverage probabilities compared to SIC for both change in mean and change in variance in the given scenarios. The impact of changing the parameters (μ, λ) is evident in the variations in coverage probabilities. As with the previous simulations, as the sample size (n) increases, both MIC and SIC show an improvement in coverage probabilities, indicating better performance with larger sample sizes.

Table 3.5: Coverage Probabilities for MIC and SIC with Change in Mean for Various Sample Sizes and Different Change Point Locations

			$(1.25, 1)$		$(1.5, 1)$		$(1.75, 1)$		$(2, 1)$	
n	k	α	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.95	0.912	0.871	0.921	0.896	0.934	0.902	0.935	0.908
		0.99	0.949	0.908	0.950	0.925	0.960	0.928	0.962	0.935
	20	0.95	0.918	0.876	0.924	0.886	0.935	0.904	0.937	0.909
		0.99	0.948	0.907	0.955	0.921	0.961	0.931	0.965	0.938
	25	0.95	0.917	0.879	0.928	0.896	0.934	0.909	0.948	0.926
		0.99	0.952	0.912	0.96	0.928	0.964	0.940	0.971	0.950
100	25	0.95	0.906	0.867	0.927	0.89	0.945	0.913	0.947	0.929
		0.99	0.951	0.911	0.967	0.928	0.976	0.945	0.980	0.962
	40	0.95	0.912	0.866	0.926	0.896	0.939	0.917	0.945	0.926
		0.99	0.966	0.921	0.971	0.941	0.975	0.953	0.98	0.963
	50	0.95	0.915	0.869	0.940	0.913	0.946	0.926	0.943	0.933
		0.99	0.964	0.919	0.969	0.944	0.974	0.954	0.972	0.962
200	50	0.95	0.906	0.866	0.937	0.907	0.949	0.930	0.950	0.939
		0.99	0.964	0.928	0.971	0.944	0.978	0.961	0.98	0.968
	75	0.95	0.906	0.873	0.944	0.919	0.961	0.943	0.957	0.950
		0.99	0.961	0.932	0.981	0.958	0.989	0.972	0.988	0.980
	100	0.95	0.905	0.875	0.95	0.919	0.953	0.935	0.955	0.954
		0.99	0.963	0.934	0.983	0.955	0.986	0.970	0.983	0.981

Table 3.6: Coverage Probabilities for MIC and SIC with Change in Variance for Various Sample Sizes and Different Change Point Locations

			$(1, 1.25)$		$(1, 1.5)$		$(1, 1.75)$		$(1, 2)$	
n	k	α	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.95	0.904	0.852	0.911	0.863	0.917	0.876	0.927	0.893
		0.99	0.947	0.893	0.952	0.901	0.956	0.916	0.962	0.926
	20	0.95	0.914	0.857	0.924	0.879	0.931	0.892	0.940	0.906
		0.99	0.945	0.887	0.951	0.905	0.961	0.918	0.964	0.927
	25	0.95	0.918	0.865	0.928	0.883	0.934	0.888	0.944	0.900
		0.99	0.947	0.894	0.954	0.909	0.961	0.918	0.970	0.928
100	25	0.95	0.895	0.853	0.911	0.875	0.922	0.893	0.936	0.912
		0.99	0.950	0.906	0.956	0.921	0.961	0.931	0.970	0.943
	40	0.95	0.895	0.857	0.918	0.880	0.938	0.906	0.946	0.923
		0.99	0.957	0.914	0.967	0.928	0.973	0.944	0.980	0.959
	50	0.95	0.901	0.861	0.914	0.875	0.939	0.914	0.948	0.927
		0.99	0.955	0.912	0.968	0.930	0.972	0.947	0.978	0.958
200	50	0.95	0.885	0.851	0.913	0.886	0.930	0.912	0.945	0.936
		0.99	0.964	0.929	0.976	0.946	0.977	0.959	0.982	0.972
	75	0.95	0.894	0.865	0.918	0.883	0.941	0.924	0.946	0.947
		0.99	0.961	0.929	0.976	0.942	0.981	0.962	0.986	0.973
	100	0.95	0.904	0.818	0.930	0.866	0.943	0.909	0.952	0.932
		0.99	0.964	0.912	0.970	0.932	0.980	0.954	0.978	0.974

Table 3.7: Coverage Probabilities for MIC and SIC with Change in Mean and Variance for Various Sample Sizes and Different Change Point Locations

			$(1.25, 1.25)$		$(1.5, 1.5)$		$(1.75, 1.75)$		$(2, 2)$	
n	k	α	MIC	SIC	MIC	SIC	MIC	SIC	MIC	SIC
50	15	0.95	0.911	0.864	0.919	0.885	0.933	0.919	0.938	0.935
		0.99	0.954	0.902	0.956	0.917	0.959	0.942	0.963	0.958
	20	0.95	0.919	0.879	0.927	0.910	0.933	0.932	0.932	0.946
		0.99	0.952	0.920	0.961	0.944	0.963	0.957	0.964	0.973
	25	0.95	0.924	0.866	0.933	0.894	0.937	0.922	0.946	0.933
		0.99	0.955	0.903	0.962	0.932	0.965	0.952	0.973	0.966
100	25	0.95	0.904	0.868	0.926	0.905	0.939	0.934	0.931	0.954
		0.99	0.957	0.917	0.968	0.948	0.962	0.973	0.960	0.982
	40	0.95	0.913	0.863	0.930	0.920	0.943	0.940	0.924	0.953
		0.99	0.966	0.917	0.973	0.954	0.972	0.973	0.969	0.984
	50	0.95	0.912	0.865	0.935	0.911	0.945	0.938	0.935	0.954
		0.99	0.963	0.913	0.966	0.946	0.974	0.963	0.971	0.974
200	50	0.95	0.903	0.868	0.943	0.938	0.945	0.971	0.935	0.969
		0.99	0.965	0.937	0.978	0.976	0.979	0.992	0.966	0.995
	75	0.95	0.915	0.874	0.952	0.948	0.953	0.968	0.946	0.969
		0.99	0.971	0.930	0.984	0.973	0.984	0.991	0.980	0.995
	100	0.95	0.934	0.882	0.964	0.950	0.963	0.968	0.949	0.976
		0.99	0.974	0.939	0.982	0.981	0.984	0.993	0.973	0.994

Chapter 4

Real Data Application

4.1 Compressive Strength and Strain of Maize Seeds

Mechanical damage during seed harvesting affects quality, so understanding strain and strength helps minimize damage during processing. This data comes from the strength data set within the goodness of fit package [SC15]. The following data was pulled from a study that found that strain and strength vary across maize genotypes and moisture levels. The strain column is a numeric vector giving the relative change in length under compression stress in millimeters. The cstrength column is a numeric vector giving the compressive strength in Newtons.

	strain	cstrength		strain	cstrength		strain	cstrength
1	0.293	245.247	31	0.226	179.807	61	0.145	131.963
2	0.274	218.23332	32	0.331	609.164	62	0.225	285.061
3	0.280	352.784	33	0.267	492.030	63	0.210	115.964
4	0.262	284.437	34	0.231	360.186	64	0.130	145.347
5	0.270	403.050	35	0.329	681.887	65	0.103	67.713
6	0.284	119.906	36	0.246	469.593	66	0.232	130.637
7	0.177	130.879	37	0.465	169.768	67	0.257	154.195
8	0.463	240.787	38	0.168	238.542	68	0.099	59.958
9	0.257	139.557	39	0.164	77.385	69	0.249	121.221
10	0.212	86.950	40	0.170	70.000	70	0.116	70.256
11	0.192	123.968	41	0.193	88.121	71	0.183	190.851
12	0.331	151.725	42	0.130	116.792	72	0.322	242.994
13	0.400	220.658	43	0.165	102.045	73	0.186	207.973
14	0.213	173.919	44	0.215	242.737	74	0.322	399.324
15	0.196	133.445	45	0.192	53.362	75	0.201	323.350
16	0.336	163.574	46	0.270	135.443	76	0.355	732.428

17	0.220	505.008	47	0.242	201.819	77	0.147	74.932
18	0.371	163.154	48	0.369	96.656	78	0.128	133.038
19	0.208	373.503	49	0.242	101.833	79	0.193	139.973
20	0.287	136.247	50	0.206	116.738	80	0.237	180.017
21	0.343	284.86	51	0.227	617.487	81	0.128	94.293
22	0.261	160.633	52	0.226	210.598	82	0.186	167.844
23	0.246	500.193	53	0.307	266.626	83	0.448	422.769
24	0.276	267.68	54	0.325	267.502	84	0.160	228.962
25	0.262	98.89	55	0.166	220.042	85	0.282	198.267
26	0.269	357.222	56	0.118	110.715	86	0.197	101.210
27	0.234	144.604	57	0.318	281.564	87	0.185	116.140
28	0.272	453.428	58	0.147	207.570	88	0.415	733.765
29	0.386	157.872	59	0.126	209.626	89	0.223	257.975
30	0.213	244.917	60	0.357	300.310	90	0.287	71.767

We will use the goodness of fit test for the Inverse Gaussian (IG) distribution which transforms the observations into approximately normally distributed observations and then use the Shapiro-Wilk test for assessing univariate normality. This dataset produces a p-value of 0.6909 meaning there is not enough evidence to reject the null hypothesis. Thus, our dataset does not significantly deviate from an IG distribution when transformed to normality.

This dataset follows the IG model, Figure 4.1 displays the change point detection results using the Modified Information Criterion (MIC) for the compressive strength of maize. To find all the change points in the dataset, we employ the binary segmentation procedure. Once we identify the location of the first change point, we split the data into two segments and apply the proposed method to each segment. This process continues until no more change points are found.

It is seen that the change points occur at positions 15, 25, 39, 50, 62, 72, and 77. This means there were significant changes in the compressive strength at those locations. Thus, we are able to more specifically pinpoint specific locations where the compressive strength of maize seeds undergo notable shifts, allowing us to dive further into the factors that may contribute to these changes.

Figure 4.2 shows the change point detection results using the Schwartz Information Criterion (SIC) for the compressive strength of maize. Change points occur at positions 16, 39, 62, 70, and 77. There is some overlap in the change points identified by MIC and SIC (both include 39 and 62). At these locations, there may be more significant changes in compressive strength.

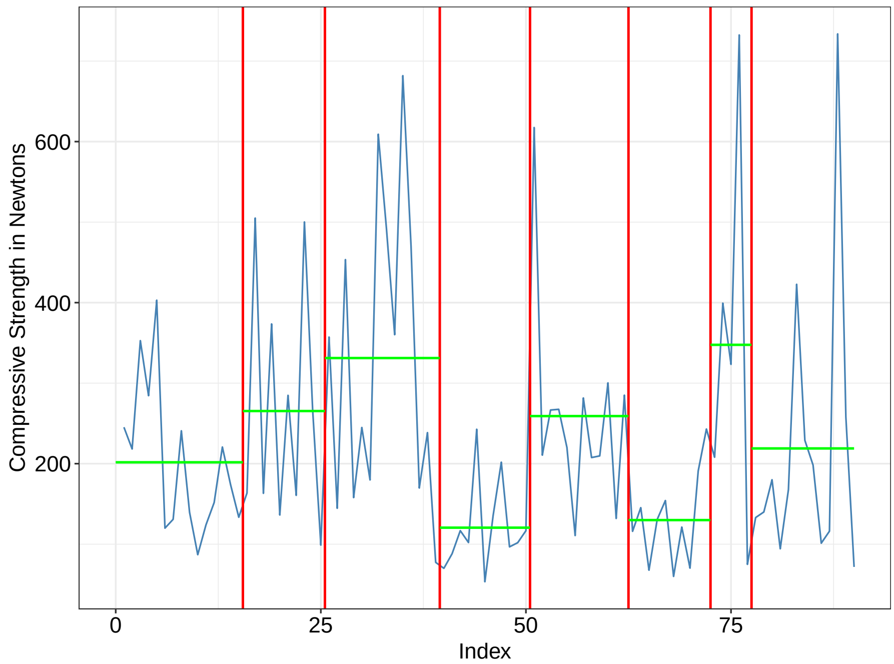

Figure 4.1: Change point detection for the Compressive Strength of Maize Seeds Using MIC

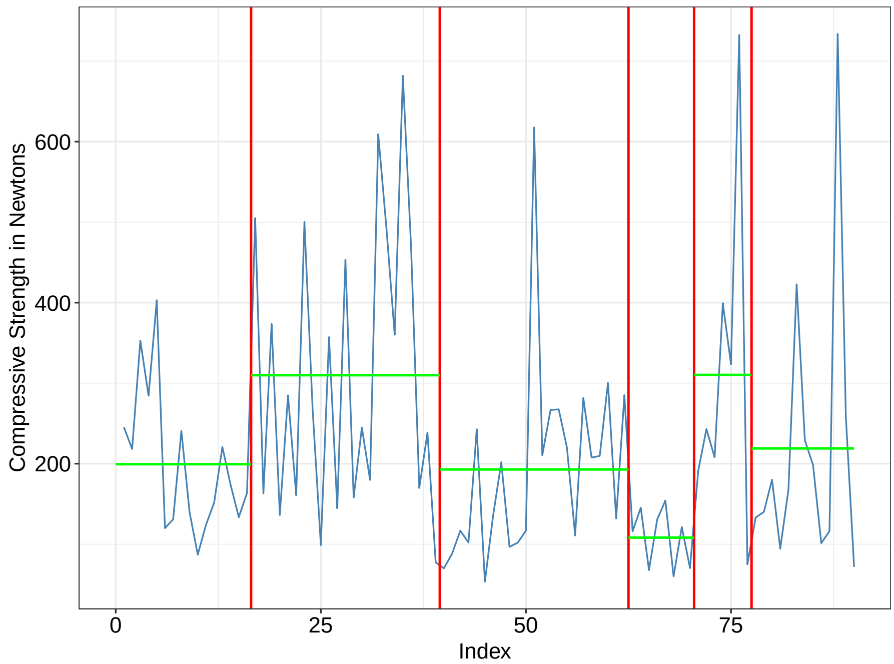

Figure 4.2: Change point detection for the Compressive Strength of Maize Seeds Using SIC

4.2 Measurement of Dispersion of Particulate Matter 2.5 in Thailand

Thailand has a high amount of fine particles less than 2.5 micrometers in diameter, also known as particulate matter $\leq 2.5\mu$m (PM 2.5). PM 2.5 contains microscopic solids or liquid droplets that are very small that they can be inhaled and cause serious health problems which is a growing concern for the public in Thailand. Large sources of PM 2.5 emissions in Thailand are transportation, coal-fired power plants, and agricultural and garbage burning. The following data comes from [C+22] which measures the daily PM 2.5 of Din Daeng district of Bangkok, Thailand in January 2021. Note that the Environmental Protection Agency states that the 24-hour fine particle standard of PM 2.5 should be 35 μm.

Date	PM 2.5	Date	PM 2.5	Date	PM 2.5
1	27.54	11	35.75	21	82.08
2	32.58	12	30	22	91.88

3	47.29	13	52.92	23	89.71
4	54.88	14	77	24	47.96
5	42.25	15	93.46	25	44.83
6	47.46	16	105.79	26	41.33
7	37.83	17	60.46	27	49.46
8	29.88	18	34.63	28	51
9	23.58	19	53.96	29	49.04
10	32.63	20	77.79	30	51.58
				31	50.26

Chankham, Wasana et al. [C^+22] uses the inverse Gaussian distribution on this data set as the inverse Gaussian distribution is good for examining the frequency of high-concentration events. Thus, the inverse Gaussian distribution is often applied to datasets like this which study air pollution. Chankham, Wasana et al. [C^+22] calculated the quantitative difference in PM 2.5 concentrations using the coefficient of variation of the Inverse Gaussian (IG) distribution. However, we will apply our MIC-based change point detection method here to see where changes occur. To check, the data was run through the goodness of fit test for the IG distribution. This data set produces a p-value of 0.2371 meaning there is not enough evidence to reject the null hypothesis. This data does not significantly deviate from an IG distribution when transformed to normality which is what we would expect as the IG distribution is often used to analyze pollution data like this.

Figure 4.3 displays the change point detection results using the Modified Information Criterion (MIC) for the measurement of dispersion of PM 2.5 in Thailand. To find all the change points in the dataset, we employ the binary segmentation procedure. Once we identify the location of the first change point, we split the data into two segments and apply the proposed method to each segment. We continue this process until no more change points are found.

After completing the binary segmentation procedure, we find that there are change points at positions 8, 11, 18, and 24. This means that at these locations there were significant changes in the dispersion of PM 2.5. Knowing this information can allow us to better specify locations where the levels of PM 2.5 are notably different, allowing statisticians to look deeper into factors that could contribute to these changes.

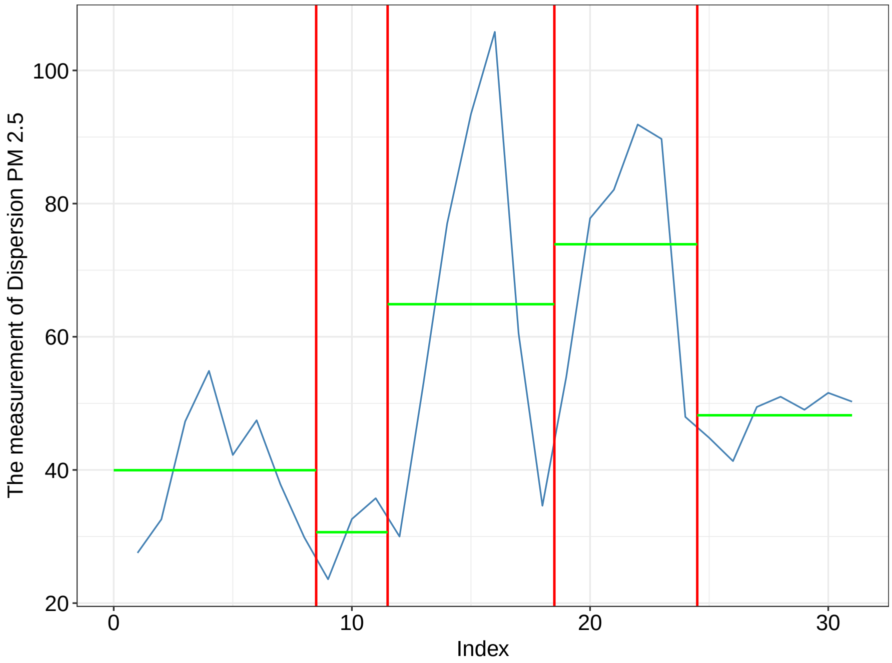

Figure 4.3: Change point detection for the measurement of Dispersion PM 2.5 Using MIC

Chapter 5

Conclusion

In this thesis, we propose a change point detection method for a two-parameter Inverse Gaussian (IG) distribution based on the Modified Information Criterion (MIC). Change point analysis aims to find the number of and corresponding location of changes in data if any exist. A simulation study is conducted to evaluate the effectiveness of the proposed detection procedure in the IG model using three criteria: type I error rate, power, and coverage probabilities.

For small n, the MIC-based procedure provides a type I error closer to the nominal level and better power compared to the SIC-based procedure. However, as the sample size increases, the type I error based on both SIC and MIC approaches closer to the nominal level, $\alpha = 0.05$. Furthermore, the power also increases; however, the MIC-based procedure outperforms the SIC-based procedure. It is clear that the power based on the SIC method is small when the true change point location k is near the beginning of the data, while the MIC-based procedure provides better power as it considers the impact of the change location in terms of model complexity. Regardless of the procedure, it is evident that power tends to increase as the difference between the pre-change and post-change parameters increases. So, for small sample sizes ($n \leq 50$), we recommend a MIC-based procedure for detecting changes in the IG model. The proposed method, along with the Binary segmentation method, is then applied to detect multiple changes in two real datasets: the compressive strength and strain of maize seeds and the measurement of dispersion of particulate matter 2.5 in Thailand.

In future research, we will explore change point analysis of the IG model applied to censored data sets, focusing on survival analysis. Further, we aim to investigate the impact of length-biased sampling on survival analysis within this framework. This research has the potential to develop new methodological approaches in survival analysis with interdisciplinary applications.

Bibliography

[C+22] Wasana Chankham et al. Measurement of dispersion of PM 2.5 in thailand using confidence intervals for the coefficient of variation of an inverse gaussian distribution. *PeerJ*, 10:e12988, 2 2022.

[CF77] R. S. Chhikara and J. L. Folks. The inverse gaussian distribution as a lifetime model. *Technometrics*, 19(4):461, Nov 1977.

[CG12] J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis with Applications to Genetics, Medicine, and Finance.* 2nd ed. Boston: Birkhauser, 2012.

[CGH18] C. Celine, H. Gudmund, and N.L. Hjort. Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference*, 195:14–34, 2018.

[CGP06] J. Chen, A. K. Gupta, and J. Pan. Information criterion and change point problem for regular models. *The Indian Journal of Statistics*, 68(2):252–282, 2006.

[CHH18] C. Cunen, G. Hermansen, and N. L. Hjort. Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference*, 195:14–34, 2018.

[CZ64] Herman Chernoff and Shelemyahu Zacks. Estimating the current mean of a normal distribution which is subject to changes in time. *The Annals of Mathematical Statistics*, 35, 09 1964.

[Pag54] E. S. Page. Continue inspection schemes. *Biometrika*, 41:100–135, 1954.

[Pag55] E. S. Page. A test for a chance in a parameter occurring at an unknown point. *Biometrika*, 42:523–527, 1955.

[Pat16] Debdeep Pati, Feb 2016.

[SC15] A. Stepanov and B. Coburn. *goft: R: Tests for the Inverse Gaussian Distributions*, 2015. R package goft version 1.3.6.

[Sch78] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[Shr23] Jacob Shreffler. Type i and type ii errors and statistical power, Mar 2023.

[SI93] Shunsuke Sato and Junko Inoue. Inverse gaussian distribution and its applications. *Denshi Joho Tsushin Gakkai Ronbunshi*, 76-A(8), Aug 1993.

[SS75a] A. K. Sen and M. S. Srivastava. On tests for detecting change in mean. *Annals of Statistics*, 3:98–108, 1975.

[SS75b] A. K. Sen and M. S. Srivastava. Some one-sided tests on change in level. *Technometrics*, 17:61–64, 1975.

[SS99] SRJ and V. Seshadri. The inverse gaussian distribution: Statistical theory and applications. *Journal of the American Statistical Association*, 94(447):990, Sep 1999.

[SW86] M. S. Srivastava and K. J. Worsley. Likelihood ratio tests for a change in the multivariate normal mean. *Journal of the American Statistical Association*, 81(393):199, Mar 1986.

[Twe57] M. C. K. Tweedie. statlstical properties of inverse gaussian distributions. *Annals of Mathematical Statistics*, 28:362–377, 1957.

[Vos81] L. J. Vostrikova. Detecting “disorder” in multidimensional random processes. *Soviet Mathematics Doklady*, 24:55–59, 1981.

[ZS07] N. Zhang and D. Siegmund. A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63:22–32, 2007.