
This version is available at https://strathprints.strath.ac.uk/58769/ 

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk

The Strathprints institutional repository (https://strathprints.strath.ac.uk) is a digital archive of University of Strathclyde research outputs. It has been developed to disseminate open access research outputs, expose data about those outputs, and enable the management and persistent access to Strathclyde's intellectual output.
Model Selection with Application to Gamma Process and Inverse Gaussian Process

M. Zhang & M. Revie
Department of Management Science
University of Strathclyde, Glasgow, UK

ABSTRACT: The gamma process and the inverse Gaussian process are widely used in condition-based maintenance. Both are suitable for modelling monotonically increasing degradation processes. One challenge for practitioners is determining which of the two processes is most appropriate in light of a real data set. A common practice is to select the one with a larger maximized likelihood. However, due to variations in the data, the maximized likelihood of the “wrong” model could be larger than that of the “right” model. This paper proposes an efficient and broadly applicable test statistic for model selection. The construction of the test statistic is based on the Fisher information. Extensive numerical study is conducted to indicate the conditions under which the gamma process can be well approximated by the inverse Gaussian process, or the other way around.

1 INTRODUCTION

The gamma process and the inverse Gaussian process were proposed by Dufresne et al. (1991) and Wasan (1968), respectively. Dufresne et al. (1991) proved that a gamma process is a limit of compound Poisson processes. They also constructed an inverse Gaussian process from compound Poisson processes. The gamma and inverse Gaussian processes are widely used for modeling degradation data. Various maintenance strategies developed in the literature when degradation data are modelled by the gamma process can be found in Al Labadi and Zarepour (2014) and Zhu et al. (2015). van Noortwijk (2009) provided an excellent review on the gamma process. Applications and generalizations of the inverse Gaussian process from compound Poisson processes. The mean and variance of the gamma distribution are \( \alpha \beta \) and \( \alpha \beta^2 \), respectively. A stochastic process \( \{X(t), t \geq 0\} \) is a gamma process if

- non-overlapping increments are independent;
- \( \forall t > s \geq 0 \), the random increment \( X(t) - X(s) \) has the gamma distribution \( Ga(\alpha(t-s), \beta) \).

The marginal distribution of the gamma process \( \{X(t), t \geq 0\} \) at time \( t \) is the gamma distribution \( Ga(\alpha t, \beta) \). \( \{X(t), t \geq 0\} \) is a stationary process having mutually independent, stationary and non-negative increments.

The inverse Gaussian distribution with mean \( u > 0 \) and shape parameter \( \lambda > 0 \), denoted by \( IG(u, \lambda) \), has probability density function

\[
 f_{IG}(x; u, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{(x-u)^2}{2u^2}\right), \quad x > 0,
\]

and cumulative distribution function

\[
 F_{IG}(x; u, \lambda) = \exp(2\lambda/u) \Phi\left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{u} - 1\right)\right) + \Phi\left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{u} - 1\right)\right), \quad x > 0.
\]

Here \( \Phi(\cdot) \) is the standard normal cumulative distribution function. The variance of the inverse Gaussian distribution is \( u^3/\lambda \). A stochastic process \( \{Y(t), t \geq 0\} \) is an inverse Gaussian process if

Here \( \gamma(\alpha, \beta x) \) is the lower incomplete gamma function. The mean and variance of the gamma distribution are \( \alpha \beta^{-1} \) and \( \alpha \beta^{-2} \), respectively. A stochastic process \( \{X(t), t \geq 0\} \) is a gamma process if

- non-overlapping increments are independent;
- \( \forall t > s \geq 0 \), the random increment \( X(t) - X(s) \) has the gamma distribution \( Ga(\alpha(t-s), \beta) \).

Mathematically, the gamma distribution with shape parameter \( \alpha > 0 \) and scale parameter \( \beta > 0 \), denoted by \( Ga(\alpha, \beta) \), has probability density function

\[
 f_{Ga}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0.
\]

Here \( \Gamma(\alpha) \) is the gamma function evaluated at \( \alpha \). The cumulative distribution function is the regularized gamma function:

\[
 F_{Ga}(x; \alpha, \beta) = \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) \, dy
 = \gamma(\alpha, \beta x) / \Gamma(\alpha), \quad x > 0.
\]
• non-overlapping increments are independent;
• ∀ t > s ≥ 0, the random increment \( Y(t) - Y(s) \) has the inverse Gaussian distribution \( IG(u(t - s), \lambda(t - s)^2) \).

Therefore, the marginal distribution of the inverse Gaussian process \( \{ Y(t), t \geq 0 \} \) at time \( t \) is the inverse Gaussian distribution \( IG(u, \lambda^2) \). \( \{ Y(t), t \geq 0 \} \) is a stationary process of which the increments are mutually independent, stationary and non-negative.

Both the gamma process and the inverse Gaussian process are suitable for modeling gradual damage introduced by continuous use. Therefore, given degradation data, the uppermost problem is selecting between the two processes the “right” model. A common practice is to select the one with a larger maximized likelihood. However, due to the variation in the data, the maximized likelihood of the “wrong” model is probably larger than that of the “right” model. In order to reduce the probability of selecting a wrong model, the data size should be sufficiently large. This paper proposes an efficient and broadly applicable test statistic for model selection by quoting the theorems in White (1982). The problem of model selection or model-misspecification detection has received much attention. A representative sample of works on model selection or model-misspecification detection has received much attention. A representative sample of works on model selection or model-misspecification detection includes Hyodo et al. (2012), Tsai et al. (2011) and Zhou et al. (2012). Note that, to select a model for the underlying degradation process is essentially to select a distribution for the degradation increments. Therefore, in what follows we focus on selecting between the gamma distribution and the inverse Gaussian distribution the right one for collected degradation data.

The remainder of the paper is organised as follows. Section 2 derives a general expression of the test statistic when the underlying stochastic law is unspecified. Section 3 conducts extensive numerical study to demonstrate the efficiency of the test statistic. Conclusions are outlined in Section 4.

2 A GENERAL FRAMEWORK

Due to the lack of space, we explain our idea by taking the gamma process as an example. For the inverse Gaussian process, the appropriate translations are obvious.

We might assume that the underlying degradation process is stationary. Hence both the inverse Gaussian process and the gamma process are suitable for fitting the degradation measurements. The following data-collecting scheme will be adopted. Let \( X(t) \) denote the degradation of a target device measured at time \( t \), with \( X(0) = 0 \). The degradation of the device is measured every \( \Delta(>0) \) units of time. The data-collecting scheme is terminated at time \( n\Delta \), \( n = 1, 2, \ldots \). Denote the collected degradation data by \( X_n = \{ x_1, x_2, \ldots, x_n \} \) in which \( x_i = X(i\Delta) - X((i - 1)\Delta), i = 1, 2, \ldots, n \). The independent random increments \( \{ x_1, x_2, \ldots, x_n \} \) have the same distribution function, denoted by \( G(x) \), \( x > 0 \). \( G(x) \) is the unknown underlying stochastic law. Let \( g(x) \) denote the corresponding probability density function. We below approximate \( G(x) \) by the gamma distribution, \( Ga(\alpha\Delta, \beta) \). To simplify the notation which follows, define two vectors of parameters: \( \theta = (\theta_1, \theta_2) = (\alpha\Delta, \beta) \) and \( \vartheta = (\vartheta_1, \vartheta_2) = (u\Delta, \lambda\Delta^2) \).

Given the degradation data \( X_n \), consider a quasi log-likelihood function of \( \theta \):

\[
\ell_{Ga}(\theta; X_n) = \frac{1}{n} \sum_{k=1}^{n} \log(f_{Ga}(x_k; \theta))
\]

\[
= n^{-1} \sum_{k=1}^{n} \left[ (\theta_1 - 1) \log(x_k) - \theta_2 x_k \right] + \theta_1 \log(\theta_2) - \log(\Gamma(\theta_1)).
\]

Find a value of \( \theta \) that maximizes \( \ell_{Ga}(\theta; X_n) \). Denote the maximizer by \( \hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2) \). \( \hat{\theta} \) is termed as the quasi maximum-likelihood (QML) estimator for \( \theta \). Notably, given the data \( X_n, n\ell_{Ga}(\theta; X_n) = \sum_{k=1}^{n} \log(f_{Ga}(x_k; \theta)) \) is the log-likelihood function of \( \theta \). Hence, the QML estimator \( \hat{\theta} \) is indeed the maximum-likelihood estimator for \( \theta \). The maximized quasi log-likelihood \( \ell_{Ga}(\hat{\theta}; X_n) \) is exactly \( 1/n \) of the maximized log-likelihood. The QML estimator for \( \theta_2 \) has a closed form: \( \hat{\theta}_2 = n\hat{\theta}_1/ \sum_{k=1}^{n} x_k \). The QML estimator for \( \theta_1, i.e. \hat{\theta}_1 \), is the solution of

\[
\log(\theta_1) - \psi(\theta_1) = \log \left( \frac{1}{n} \sum_{k=1}^{n} x_k \right) - \frac{1}{n} \sum_{k=1}^{n} \log(x_k),
\]

which can be solved numerically. Here, \( \psi(\theta_1) \) is the digamma function evaluated at \( \theta_1 \).

The Kullback-Leibler divergence (Kullback and Leibler 1951) is a non-symmetric measure of the difference between two probability distributions. Assume that \( G_1 \) and \( G_2 \) are two probability measures over a set \( S \), and \( G_1 \) is absolutely continuous with respect to \( G_2 \). The measure is non-symmetric in that the Kullback-Leibler divergence of \( G_1 \) from \( G_2 \) is most often different from the Kullback-Leibler divergence of \( G_2 \) from \( G_1 \). The Kullback-Leibler divergence of \( G_2 \) from \( G_1 \) is defined to be

\[
I(G_2, G_1) = \int_S \log(dG_1/dG_2) dG_1.
\]

d\( G_1/dG_2 \) is the Radon-Nikodym derivative of \( G_1 \) with respect to \( G_2 \). \( I(G_2, G_1) \) measures the information lost when \( G_2 \) (typically, a specified probability distribution) is used to approximate \( G_1 \) (typically, the true stochastic law).
The Kullback-Leibler divergence of $F_{Ga}(x; \theta)$ from $G(x)$ is
\[
I(F_{Ga}, G) = E[\log(g(x)/f_{Ga}(x; \theta))] = E[\log(g(x))] - E[\log(f_{Ga}(x; \theta))].
\]
Here, and in what follows, expectations are all taken with respect to the true stochastic law $g(x)$. The first term of the right-hand side, i.e. $E[\log(g(x))]$, is independent of $\theta$. The minimization of $I(F_{Ga}, G)$ is equivalent to the maximization of $E[\log(f_{Ga}(x; \theta))]$. Notably, the quasi log-likelihood $\ell_{Ga}(\theta; X_n)$ is a (strongly) consistent estimator for $E[\log(f_{Ga}(x; \theta))]$. Let $\theta^*$ denote the optimal parameter vector minimizing the Kullback-Leibler divergence:
\[
\theta^* = \arg \min_{\theta>0} I(F_{Ga}, G) = \arg \max_{\theta>0} E[\log(f_{Ga}(x; \theta))].
\]
If the stochastic law lies within the family of gamma distributions (i.e., $g(x) = f_{Ga}(x; \theta^*)$ for some $\theta^* > 0$), then $I(F_{Ga}, G)$ attains its unique minimum at $\theta^* = \theta_0$. The value of $\theta^*$ is inaccessible. Because $\ell_{Ga}(\theta; X_n)$ is a consistent estimator for $E[\log(f_{Ga}(x; \theta))]$, one may conjecture that $\hat{\theta}$ is a consistent estimator for $\theta^*$.

Define two $2 \times 2$ matrices $A_{Ga}(\theta)$ and $B_{Ga}(\theta)$:
\[
[A_{Ga}(\theta)]_{ij} = E\left[\frac{\partial^2 \log(f_{Ga}(x; \theta))}{\partial \theta_i \partial \theta_j}\right],
\]
and
\[
[B_{Ga}(\theta)]_{ij} = E\left[\frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_i} \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_j}\right].
\]
Here, we have utilized the fact that differentiation can be taken inside integral. $A_{Ga}(\theta)$ and $B_{Ga}(\theta)$ can be respectively consistently estimated by matrices $A_{Ga}^n(\theta)$ and $B_{Ga}^n(\theta)$:
\[
[A_{Ga}^n(\theta)]_{ij} = n^{-1} \sum_{k=1}^n \frac{\partial^2 \log(f_{Ga}(x_k; \theta))}{\partial \theta_i \partial \theta_j},
\]
and
\[
[B_{Ga}^n(\theta)]_{ij} = n^{-1} \sum_{k=1}^n \frac{\partial \log(f_{Ga}(x_k; \theta))}{\partial \theta_i} \frac{\partial \log(f_{Ga}(x_k; \theta))}{\partial \theta_j}.
\]
If the matrices $A_{Ga}(\theta)$ and $A_{Ga}^n(\theta)$ are invertible, which can always be guaranteed, define
\[
C_{Ga}(\theta) = A_{Ga}(\theta)^{-1} B_{Ga}(\theta) A_{Ga}(\theta)^{-1},
\]
and
\[
C_{Ga}^n(\theta) = A_{Ga}^n(\theta)^{-1} B_{Ga}^n(\theta) A_{Ga}^n(\theta)^{-1}.
\]
The superscript “$-1$” above a matrix denotes the inverse operator. $C_{Ga}(\theta)$ is a consistent estimator for $C_{Ga}(\theta)$.

**Proposition 1** The distribution of $\sqrt{n}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean zero and covariance matrix $C_{Ga}(\theta^*)$. The sequence $C_{Ga}^n(\theta) \xrightarrow{a.s.} C_{Ga}(\theta^*)$, element by element. Specifically, if $g(x) = f_{Ga}(x; \theta^*)$ for some $\theta^* > 0$, then

- $\hat{\theta}$ is a (strongly) consistent estimator for $\theta^*$;
- $\sqrt{n}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean zero and covariance matrix $C_{Ga}(\theta^*)$.

**Proof.** The assumptions A1-A6 in White (1982) all hold. The proposition follows from Theorem 3.2 of White (1982), and the proof is complete.

If the underlying stochastic law is correctly specified and if differentiation can be taken inside integral, the information matrix can be written in either the Hessian form, i.e. $-A_{Ga}(\theta^*)$, or the outer product form, i.e. $B_{Ga}(\theta^*)$. The information-matrix equivalence indicates that the sum $A_{Ga}(\theta^*) + B_{Ga}(\theta^*)$ can be used for detecting model misspecification. Specifically, the failure of the sum $A_{Ga}(\theta^*) + B_{Ga}(\theta^*)$ equalling zero states that the stochastic law is misspecified. The values of the elements in $A_{Ga}(\theta^*) + B_{Ga}(\theta^*)$ are inaccessible. Yet, by Proposition 1, $A_{Ga}(\theta^*) + B_{Ga}(\theta^*)$ can be consistently estimated by $A_{Ga}^n(\theta) + B_{Ga}^n(\theta)$. Hence, the remaining work is to investigate the distributional property of the elements in $A_{Ga}^n(\theta) + B_{Ga}^n(\theta)$.

Define a vector-valued function: $\delta_{Ga}(x; \theta) = (\delta_1(x; \theta), \delta_2(x; \theta), \delta_3(x; \theta))^t$ in which
\[
\delta_1(x; \theta) = \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_1} \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_1} + \frac{\partial^2 \log(f_{Ga}(x; \theta))}{\partial \theta_1 \partial \theta_1},
\]
\[
\delta_2(x; \theta) = \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_1} \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_2} + \frac{\partial^2 \log(f_{Ga}(x; \theta))}{\partial \theta_1 \partial \theta_2},
\]
and
\[
\delta_3(x; \theta) = \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_2} \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_2} + \frac{\partial^2 \log(f_{Ga}(x; \theta))}{\partial \theta_2 \partial \theta_2}.
\]
By the superscript “$t$”, we mean the transpose of a vector or a matrix. Define
\[
\delta_{Ga}^n(x; \theta) = n^{-1} \sum_{k=1}^n \delta_{Ga}(x_k; \theta).
\]
\( \delta_{Ga}(\theta) \) consists of all the distinct elements in \( A_{Ga}^n(\theta) + B_{Ga}^n(\theta) \). Hence, we below investigate the asymptotic joint distribution of \( \delta_{Ga}^n(\theta) \). Take expectation of \( \delta_{Ga}(x; \theta) \) with respect to \( g(x) \):

\[
\delta_{Ga}(\theta) = E[\delta_{Ga}(x; \theta)] = (E[\delta_1(x; \theta)], E[\delta_2(x; \theta)], E[\delta_3(x; \theta)])^t.
\]

The respective \( 3 \times 2 \) Jacobian matrices of the vector-valued functions \( \delta_{Ga}(\theta) \) and \( \delta_{Ga}(\theta) \) are

\[
[J_{Ga}^n(\theta)]_{ij} = n^{-1} \sum_{k=1}^n \frac{\partial \delta_i(x_k; \theta)}{\partial \theta_j},
\]

and

\[
[J_{Ga}(\theta)]_{ij} = \frac{\partial E[\delta_i(x; \theta)]}{\partial \theta_j} = E\left[ \frac{\partial \delta_i(x; \theta)}{\partial \theta_j} \right].
\]

The partial derivative with respect to \( \theta \) of the logarithm of \( f_{Ga}(x; \theta) \) is

\[
\nabla \log(f_{Ga}(x; \theta)) = \left( \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_1}, \frac{\partial \log(f_{Ga}(x; \theta))}{\partial \theta_2} \right)^t.
\]

The nabla symbol “\( \nabla \)” denotes the vector differential operator. Define two \( 3 \times 3 \) matrices

\[
V_{Ga}(\theta) = E[\vec{v}_{Ga}(x; \theta)\vec{v}_{Ga}(x; \theta)^t],
\]

and

\[
V_{Ga}^n(\theta) = n^{-1} \sum_{k=1}^n \vec{v}_{Ga}(x_k; \theta)\vec{v}_{Ga}(x_k; \theta)^t.
\]

The column vectors \( \vec{v}_{Ga}(x; \theta) \) and \( \vec{v}_{Ga}(x_k; \theta) \) are defined by

\[
\vec{v}_{Ga}(x; \theta) = \delta_{Ga}(x; \theta)
\]

\[-J_{Ga}(\theta)A_{Ga}(\theta)^{-1}\nabla \log(f_{Ga}(x; \theta)),
\]

and

\[
\vec{v}_{Ga}(x_k; \theta) = \delta_{Ga}(x_k; \theta)
\]

\[-J_{Ga}^n(\theta)A_{Ga}^n(\theta)^{-1}\nabla \log(f_{Ga}(x_k; \theta)).
\]

**Proposition 2** If \( g(x) = f_{Ga}(x; \theta^0) \) for some \( \theta^0 > 0 \), then

- \( \sqrt{n} \delta_{Ga}^n(\theta^0) \) is asymptotically normally distributed with mean zero and covariance matrix \( V_{Ga}(\theta^0) \);
- the sequence \( V_{Ga}^n(\hat{\theta}) \) converges almost surely to \( V_{Ga}(\theta^0) \);
- \( V_{Ga}^n(\hat{\theta}) \) is nonsingular almost surely for all sufficiently large \( n \);
- the asymptotic distribution of the test statistic \( \zeta_{Ga}^n = n \delta_{Ga}^n(\theta)^tV_{Ga}^n(\theta)^{-1}\delta_{Ga}^n(\theta) \) is chi-squared distribution with three degrees of freedom.

**Proof.** The assumptions A1-A10 in White (1982) all hold. The proposition follows from Theorem 4.1 of White (1982), and the proof is complete.

\( \zeta_{Ga}^n \) serves as a test statistic in a hypothesis test. A hypothesis test can be constructed with the null hypothesis given by

\[ H_0 : g(x) = f_{Ga}(x; \theta), \quad \exists \theta > 0 \]

and the alternative hypothesis given by

\[ H_1 : g(x) \neq f_{Ga}(x; \theta), \quad \forall \theta > 0. \]

If the null hypothesis is true, the test statistic \( \zeta_{Ga}^n \) is chi-squared distributed with three degrees of freedom. To carry out the test, one calculates \( \zeta_{Ga}^n \) and compares it to the critical value of the \( \chi^2_3 \) distribution. If \( \zeta_{Ga}^n \) exceeds the critical value, one rejects the null hypothesis and concludes that the specified family of probability distributions is inappropriate. If the null hypothesis is accepted, one may have confidence that the estimators will be consistent for parameters of interest. Note that \( \zeta_{Ga}^n \) also measures how well the density function \( g(x) \) could be approximated by a gamma density function. Specifically, for a given sample size \( n \), the smaller the value of \( \zeta_{Ga}^n \), the better the density function \( g(x) \) could be approximated by a gamma density function.

### 3 NUMERICAL EXAMPLES

#### 3.1 Fit Data by the Gamma Process

Assume that the underlying stochastic law is an inverse Gaussian process. Randomly simulate \( n \) observations, denoted by \( X_n \), from the inverse Gaussian distribution \( f(x; \theta) \). Fit the gamma distribution to the data \( X_n \). Maximize the quasi log-likelihood \( \ell_{Ga}(\theta; X_n) \) to obtain the QML estimate of \( \theta \), i.e. \( \hat{\theta} \). The corresponding maximized quasi log-likelihood is denoted by \( \ell_{\hat{\theta}} \). Given the data \( X_n \) and the QML estimate \( \hat{\theta} \), calculate the value of the test statistic \( \zeta_{Ga}^n \). For comparison, fit the inverse Gaussian distribution to the data \( X_n \), maximize the quasi log-likelihood \( \ell_{IG}(\theta; X_n) \), and denote the maximized quasi log-likelihood by \( \ell_{\hat{\theta}} \). Define a statistic \( \tau_n \):

\[ \tau_n = P\left( \ell_{\hat{\theta}} > \ell_{\hat{\theta}} \right) = P\left( n\ell_{\hat{\theta}} > n\ell_{\hat{\theta}} \right). \]

A common practice for model selection is comparing the maximized log-likelihoods: \( n\ell_{\hat{\theta}} \) and \( n\ell_{\hat{\theta}} \). Hence, for a given sample size \( n \), \( \tau_n \) is the probability of selecting the wrong model.
For illustrative purpose, we increase the shape parameter \( \vartheta_2 \) from 0.2 to 6 with step size 0.2. Fix the mean \( \vartheta_1 \) at 10. Gradually increase the sample size \( n \) from 10 to 100 with step size 5. For each combination of \( \vartheta_2 \) and \( n \), we generate 1000 data sets. Calculate \( f^n_\vartheta \), \( f^n_\vartheta^{\vartheta_2} \) and \( \zeta^n_{\vartheta_2} \) for each data set. For the 1000 pairs of \( \{f^n_\vartheta, f^n_\vartheta^{\vartheta_2}\} \), calculate the percentage of \( f^n_\vartheta^{\vartheta_2} \) being larger than \( f^n_\vartheta \). The percentage is an estimate of \( \tau_n \). Take average of the 1000 values of \( \zeta^n_{\vartheta_2} \). The critical value of the chi-square distribution with 3 degrees of freedom at significance level 0.05 is 7.815.

Plot the evolution of \( \tau_n \) and the averaged value of \( \zeta^n_{\vartheta_2} \) in Figure 1. In Figure 1, the red curve corresponds to \( \vartheta_2 = 0.2 \), and the green curve corresponds to \( \vartheta_2 = 6 \). Figure 1 shows that, if the shape parameter \( \vartheta_2 \) is relatively large, the inverse Gaussian distribution can be closely approximated by a gamma distribution. To guarantee a 0.95 probability of selecting the right model, \( \tau_n \) should be smaller than 0.05. We let \( n_1 \) denote the sample size, beyond which the value of \( \tau_n \) is smaller than 0.05; let \( n_2 \) denote the sample size, beyond which the averaged value of \( \zeta^n_{\vartheta_2} \) is larger than 7.815. The evolution of \( n_1 \) and \( n_2 \) along with \( \vartheta_2 \) is summarized in Table 1. Table 1 shows that selecting the model according to maximized log-likelihoods is as efficient as selecting model according to the proposed test statistic.

Table 1: The required sample size for selecting the right model with probability 0.95.

<table>
<thead>
<tr>
<th>( \vartheta_2 )</th>
<th>( n_1(\tau_n) )</th>
<th>( n_2(\zeta^n_{\vartheta_2}) )</th>
<th>( \vartheta_2 )</th>
<th>( n_1(\tau_n) )</th>
<th>( n_2(\zeta^n_{\vartheta_2}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>20</td>
<td>20</td>
<td>3.4</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>0.6</td>
<td>20</td>
<td>20</td>
<td>3.8</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>1.0</td>
<td>20</td>
<td>25</td>
<td>4.2</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>1.4</td>
<td>25</td>
<td>25</td>
<td>4.6</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>1.8</td>
<td>25</td>
<td>30</td>
<td>5.0</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>30</td>
<td>30</td>
<td>5.4</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>2.6</td>
<td>35</td>
<td>35</td>
<td>5.8</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>3.0</td>
<td>30</td>
<td>35</td>
<td>6.0</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>

3.2 Fit Data by the inverse Gaussian Process

Assume that the underlying stochastic law is a gamma process. Randomly simulate \( n \) observations from the gamma distribution \( f_\vartheta(x; \vartheta) \). Fit the inverse Gaussian distribution to the data \( X_n \). Maximize the quasi log-likelihood to obtain the QML estimate of \( \vartheta \). The corresponding maximized quasi log-likelihood is denoted by \( f^n_\vartheta^{\vartheta_2} \).

For a given sample size \( n \), \( \tau_n \) is the probability of selecting the wrong model, if we select by comparing maximized log-likelihoods.

For illustrative purpose, we increase the shape parameter \( \vartheta_1 \) from 1.8 to 7 with step size 0.2. (Experiments showed that, when \( \vartheta_1 < 1.8 \) and \( n \) is small, the matrix \( V^n_{IG}(\vartheta) \) is likely to be singular. Hence, we start from 1.8 instead of from 0.2.) Fix the scale parameter \( \vartheta_2 \) at 1. Gradually increase the sample size \( n \) from 10 to 100 with step size 5. For each combination of \( \vartheta_1 \) and \( n \), we generate 1000 data sets. Calculate \( \ell^n_\vartheta^{\vartheta_1} \), \( \ell^n_\vartheta^{\vartheta_2} \) and \( \zeta^n_{IG} \) for each data set. For the 1000 pairs of \( \{\ell^n_\vartheta^{\vartheta_1}, \ell^n_\vartheta^{\vartheta_2}\} \), calculate the percentage of \( \ell^n_\vartheta^{\vartheta_1} \) being smaller than \( \ell^n_\vartheta^{\vartheta_2} \). The percentage is an estimate of \( \tau_n \). Take average of the 1000 values of \( \zeta^n_{IG} \).

Plot the evolution of \( \tau_n \) and the averaged value of \( \zeta^n_{IG} \) in Figure 2. In Figure 2, the red curve corresponds to \( \vartheta_1 = 1.8 \), and the green curve corresponds to \( \vartheta_1 = 7 \). The upper panel in Figure 2 shows that, if the shape parameter \( \vartheta_1 \) is relatively large, the gamma distribution can be closely approximated by an inverse Gaussian distribution. We let \( n_1 \) denote the sample size, beyond which the value of \( \tau_n \) is smaller than 0.05; let \( n_2 \) denote the sample size, beyond which the averaged...
value of $\zeta_{IG}^n$ is larger than 7.815. The evolution of $n_1$ and $n_2$ along with $\theta_1$ is summarized in Table 2. From Table 2, it is clear that the proposed test statistic is much more efficient than $\tau_n$. By comparing Figures 1 and 2 and Tables 1 and 2, it can be found that

- from $\tau_n$ point of view, when fitting a gamma distribution to inverse-Gaussian distributed data, the probability of selecting the wrong model is small; when fitting an inverse Gaussian distribution to gamma distributed data, the probability of selecting the wrong model is relatively large;

- from $\zeta_{Ga}^n$ or $\zeta_{IG}^n$ point of view, when fitting a gamma distribution to inverse-Gaussian distributed data, the required sample size increases with the shape parameter $\theta_2$; when fitting an inverse Gaussian distribution to gamma distributed data, the required sample size is very small.

Because the proposed test statistic is more efficient than $\tau_n$, we say that the gamma distribution is more flexible than the inverse Gaussian distribution.

4 CONCLUSIONS

This paper proposed a test statistic for model selection (or, model-misspecification detection). The gamma process and the inverse Gaussian process were used for illustration, due to their wide applications and essential similarities. Numerical study showed that the proposed approach is more effective than selecting model based on maximized likelihoods. It was found that the inverse Gaussian density function with a large shape parameter can be well approximated by a gamma density function. The construction of the statistic is based on the Fisher information. Therefore, the statistic is broadly applicable to cases in which the following two qualifications are satisfied.

- The information matrix can be expressed in both the Hessian form and the outer product form.

- The Hessian matrix of the logarithm of the density function is invertible.

REFERENCES


---

**Table 2**: The required sample size for selecting the right model with probability 0.95.

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$n_1(\tau_n)$</th>
<th>$n_2(\zeta_{IG}^n)$</th>
<th>$\theta_1$</th>
<th>$n_1(\tau_n)$</th>
<th>$n_2(\zeta_{IG}^n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8</td>
<td>95</td>
<td>20</td>
<td>4.6</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>90</td>
<td>20</td>
<td>5.0</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>2.6</td>
<td>$&gt;100$</td>
<td>20</td>
<td>5.4</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>3.0</td>
<td>$&gt;100$</td>
<td>20</td>
<td>5.8</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>3.4</td>
<td>$&gt;100$</td>
<td>20</td>
<td>6.2</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>3.8</td>
<td>$&gt;100$</td>
<td>20</td>
<td>6.6</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
<tr>
<td>4.2</td>
<td>$&gt;100$</td>
<td>20</td>
<td>7.0</td>
<td>$&gt;100$</td>
<td>20</td>
</tr>
</tbody>
</table>