

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 Zhang and Zhou (2014), Zhang et al. (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 can be found in Al Labadi and Zarepour (2013), Griffin et al. (2013) and Tang et al. (2009).

Mathematically, the gamma distribution with shape parameter α (> 0) and scale parameter β (> 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 α . The cumulative distribution function is the regularized gamma function:

$$\begin{aligned} F_{Ga}(x; \alpha, \beta) &= \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) dy \\ &= \gamma(\alpha, \beta x) / \Gamma(\alpha), \quad x > 0. \end{aligned}$$

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 $Ga(\alpha(t-s), \beta)$.

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 λ (> 0), denoted by $IG(u, \lambda)$, has probability density function

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

and cumulative distribution function

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

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

- non-overlapping increments are independent;
- $\forall t > s \geq 0$, 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 $IG(ut, \lambda t^2)$. $\{Y(t), t \geq 0\}$ 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 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 $n\Delta$, $n = 1, 2, \dots$. Denote the collected degradation data by $X_n = \{x_1, x_2, \dots, x_n\}$ in which $x_i = X(i\Delta) -$

$X((i-1)\Delta)$, $i = 1, 2, \dots, n$. The independent random increments $\{x_1, x_2, \dots, x_n\}$ 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, $G_a(\alpha\Delta, \beta)$. 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 θ :

$$\begin{aligned} \ell_{G_a}(\theta; X_n) &= \frac{1}{n} \sum_{k=1}^n \log(f_{G_a}(x_k; \theta)) \\ &= n^{-1} \sum_{k=1}^n [(\theta_1 - 1) \log(x_k) - \theta_2 x_k] \\ &\quad + \theta_1 \log(\theta_2) - \log(\Gamma(\theta_1)). \end{aligned}$$

Find a value of θ that maximizes $\ell_{G_a}(\theta; X_n)$. Denote the maximizer by $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$. $\hat{\theta}$ is termed as the quasi maximum-likelihood (QML) estimator for θ . Notably, given the data X_n , $n\ell_{G_a}(\theta; X_n) = \sum_{k=1}^n \log(f_{G_a}(x_k; \theta))$ is the log-likelihood function of θ . Hence, the QML estimator $\hat{\theta}$ is indeed the maximum-likelihood estimator for θ . The maximized quasi log-likelihood $\ell_{G_a}(\hat{\theta}; X_n)$ is exactly $1/n$ of the maximized log-likelihood. The QML estimator for θ_2 has a closed form: $\hat{\theta}_2 = n\hat{\theta}_1 / \sum_{k=1}^n x_k$. The QML estimator for θ_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 θ_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

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

dG_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 $F_{G_a}(x; \theta)$ from $G(x)$ is

$$\begin{aligned} I(F_{G_a}, G) &= E[\log(g(x)/f_{G_a}(x; \theta))] \\ &= E[\log(g(x))] - E[\log(f_{G_a}(x; \theta))]. \end{aligned}$$

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 θ . The minimization of $I(F_{G_a}, G)$ is equivalent to the maximization of $E[\log(f_{G_a}(x; \theta))]$. Notably, the quasi log-likelihood $\ell_{G_a}(\theta; X_n)$ is a (strongly) consistent estimator for $E[\log(f_{G_a}(x; \theta))]$. Let θ^* denote the optimal parameter vector minimizing the Kullback-Leibler divergence:

$$\theta^* = \arg \min_{\theta > 0} I(F_{G_a}, G) = \arg \max_{\theta > 0} E[\log(f_{G_a}(x; \theta))].$$

If the stochastic law lies within the family of gamma distributions (i.e., $g(x) = f_{G_a}(x; \theta^0)$ for some $\theta^0 > 0$), then $I(F_{G_a}, G)$ attains its unique minimum at $\theta^* = \theta^0$. The value of θ^* is inaccessible. Because $\ell_{G_a}(\theta; X_n)$ is a consistent estimator for $E[\log(f_{G_a}(x; \theta))]$, one may conjecture that $\hat{\theta}$ is a consistent estimator for θ^* .

Define two 2×2 matrices $A_{G_a}(\theta)$ and $B_{G_a}(\theta)$:

$$[A_{G_a}(\theta)]_{ij} = E \left[\frac{\partial^2 \log(f_{G_a}(x; \theta))}{\partial \theta_i \partial \theta_j} \right],$$

and

$$[B_{G_a}(\theta)]_{ij} = E \left[\frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_i} \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_j} \right].$$

Here, we have utilized the fact that differentiation can be taken inside integral. $A_{G_a}(\theta)$ and $B_{G_a}(\theta)$ can be respectively consistently estimated by matrices $A_{G_a}^n(\theta)$ and $B_{G_a}^n(\theta)$:

$$[A_{G_a}^n(\theta)]_{ij} = n^{-1} \sum_{k=1}^n \frac{\partial^2 \log(f_{G_a}(x_k; \theta))}{\partial \theta_i \partial \theta_j},$$

and

$$[B_{G_a}^n(\theta)]_{ij} = n^{-1} \sum_{k=1}^n \frac{\partial \log(f_{G_a}(x_k; \theta))}{\partial \theta_i} \frac{\partial \log(f_{G_a}(x_k; \theta))}{\partial \theta_j}.$$

If the matrices $A_{G_a}(\theta)$ and $A_{G_a}^n(\theta)$ are invertible, which can always be guaranteed, define

$$C_{G_a}(\theta) = A_{G_a}(\theta)^{-1} B_{G_a}(\theta) A_{G_a}(\theta)^{-1},$$

and

$$C_{G_a}^n(\theta) = A_{G_a}^n(\theta)^{-1} B_{G_a}^n(\theta) A_{G_a}^n(\theta)^{-1}.$$

The superscript “-1” above a matrix denotes the inverse operator. $C_{G_a}^n(\theta)$ is a consistent estimator for $C_{G_a}(\theta)$.

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

- $\hat{\theta}$ is a (strongly) consistent estimator for θ^0 ;
- $\sqrt{n}(\hat{\theta} - \theta^0)$ is asymptotically normal with mean zero and covariance matrix $C_{G_a}(\theta^0)$.

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 expressed in either the Hessian form, i.e. $-A_{G_a}(\theta^0)$, or the outer product form, i.e. $B_{G_a}(\theta^0)$. The information-matrix equivalence indicates that the sum $A_{G_a}(\theta^0) + B_{G_a}(\theta^0)$ can be used for detecting model misspecification. Specifically, the failure of the sum $A_{G_a}(\theta^*) + B_{G_a}(\theta^*)$ equalling zero states that the stochastic law is misspecified. The values of the elements in $A_{G_a}(\theta^*) + B_{G_a}(\theta^*)$ are inaccessible. Yet, by Proposition 1, $A_{G_a}(\theta^*) + B_{G_a}(\theta^*)$ can be consistently estimated by $A_{G_a}^n(\hat{\theta}) + B_{G_a}^n(\hat{\theta})$. Hence, the remaining work is to investigate the distributional property of the elements in $A_{G_a}^n(\hat{\theta}) + B_{G_a}^n(\hat{\theta})$.

Define a vector-valued function: $\delta_{G_a}(x; \theta) = (\delta_1(x; \theta), \delta_2(x; \theta), \delta_3(x; \theta))^t$ in which

$$\begin{aligned} \delta_1(x; \theta) &= \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_1} \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_1} \\ &\quad + \frac{\partial^2 \log(f_{G_a}(x; \theta))}{\partial \theta_1 \partial \theta_1}, \end{aligned}$$

$$\begin{aligned} \delta_2(x; \theta) &= \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_1} \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_2} \\ &\quad + \frac{\partial^2 \log(f_{G_a}(x; \theta))}{\partial \theta_1 \partial \theta_2}, \end{aligned}$$

and

$$\begin{aligned} \delta_3(x; \theta) &= \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_2} \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_2} \\ &\quad + \frac{\partial^2 \log(f_{G_a}(x; \theta))}{\partial \theta_2 \partial \theta_2}. \end{aligned}$$

By the superscript “t”, we mean the transpose of a vector or a matrix. Define

$$\bar{\delta}_{G_a}^n(\theta) = n^{-1} \sum_{k=1}^n \delta_{G_a}(x_k; \theta).$$

$\bar{\delta}_{G_a}^n(\hat{\theta})$ consists of all the distinct elements in $A_{G_a}^n(\hat{\theta}) + B_{G_a}^n(\hat{\theta})$. Hence, we below investigate the asymptotic joint distribution of $\bar{\delta}_{G_a}^n(\hat{\theta})$. Take expectation of $\delta_{G_a}(x; \theta)$ with respect to $g(x)$:

$$\begin{aligned}\bar{\delta}_{G_a}(\theta) &= E[\delta_{G_a}(x; \theta)] \\ &= (E[\delta_1(x; \theta)], E[\delta_2(x; \theta)], E[\delta_3(x; \theta)])^t.\end{aligned}$$

The respective 3×2 Jacobian matrices of the vector-valued functions $\bar{\delta}_{G_a}^n(\theta)$ and $\bar{\delta}_{G_a}(\theta)$ are

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

and

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

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

$$\begin{aligned}\nabla \log(f_{G_a}(x; \theta)) \\ = \left(\frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_1}, \frac{\partial \log(f_{G_a}(x; \theta))}{\partial \theta_2} \right)^t.\end{aligned}$$

The nabla symbol “ ∇ ” denotes the vector differential operator. Define two 3×3 matrices

$$V_{G_a}(\theta) = E[\mathbf{v}_{G_a}(x; \theta) \mathbf{v}_{G_a}(x; \theta)^t],$$

and

$$V_{G_a}^n(\theta) = n^{-1} \sum_{k=1}^n \ddot{\mathbf{v}}_{G_a}(x_k; \theta) \ddot{\mathbf{v}}_{G_a}(x_k; \theta)^t.$$

The column vectors $\mathbf{v}_{G_a}(x; \theta)$ and $\ddot{\mathbf{v}}_{G_a}(x_k; \theta)$ are defined by

$$\begin{aligned}\mathbf{v}_{G_a}(x; \theta) &= \delta_{G_a}(x; \theta) \\ &\quad - J_{G_a}(\theta) A_{G_a}(\theta)^{-1} \nabla \log(f_{G_a}(x; \theta)),\end{aligned}$$

and

$$\begin{aligned}\ddot{\mathbf{v}}_{G_a}(x_k; \theta) &= \delta_{G_a}(x_k; \theta) \\ &\quad - J_{G_a}^n(\theta) A_{G_a}^n(\theta)^{-1} \nabla \log(f_{G_a}(x_k; \theta)).\end{aligned}$$

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

- $\sqrt{n} \bar{\delta}_{G_a}^n(\hat{\theta})$ is asymptotically normally distributed with mean zero and covariance matrix $V_{G_a}(\theta^0)$;
- the sequence $V_{G_a}^n(\hat{\theta})$ converges almost surely to $V_{G_a}(\theta^0)$;

- $V_{G_a}^n(\hat{\theta})$ is nonsingular almost surely for all sufficiently large n ;
- the asymptotic distribution of the test statistic $\zeta_{G_a}^n = n \bar{\delta}_{G_a}^n(\hat{\theta})^t V_{G_a}^n(\hat{\theta})^{-1} \bar{\delta}_{G_a}^n(\hat{\theta})$ 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.

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

$$H_0 : g(x) = f_{G_a}(x; \theta), \quad \exists \theta > 0$$

and the alternative hypothesis given by

$$H_1 : g(x) \neq f_{G_a}(x; \theta), \quad \forall \theta > 0.$$

If the null hypothesis is true, the test statistic $\zeta_{G_a}^n$ is chi-squared distributed with three degrees of freedom. To carry out the test, one calculates $\zeta_{G_a}^n$ and compares it to the critical value of the χ_3^2 distribution. If $\zeta_{G_a}^n$ 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 $\zeta_{G_a}^n$ 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 $\zeta_{G_a}^n$, 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 $f_{IG}(x; \vartheta)$. Fit the gamma distribution to the data X_n . Maximize the quasi log-likelihood $\ell_{G_a}(\theta; X_n)$ to obtain the QML estimate of θ , i.e. $\hat{\theta}$. The corresponding maximized quasi log-likelihood is denoted by $\ell_{\hat{\theta}}^n$. Given the data X_n and the QML estimate $\hat{\theta}$, calculate the value of the test statistic $\zeta_{G_a}^n$. For comparison, fit the inverse Gaussian distribution to the data X_n , maximize the quasi log-likelihood $\ell_{IG}(\vartheta; X_n)$, and denote the maximized quasi log-likelihood by $\ell_{\hat{\vartheta}}^n$. Define a statistic τ_n :

$$\tau_n = P(\ell_{\hat{\theta}}^n > \ell_{\hat{\vartheta}}^n) = P(n \ell_{\hat{\theta}}^n > n \ell_{\hat{\vartheta}}^n).$$

A common practice for model selection is comparing the maximized log-likelihoods: $n \ell_{\hat{\theta}}^n$ and $n \ell_{\hat{\vartheta}}^n$. Hence, for a given sample size n , τ_n is the probability of selecting the wrong model.

For illustrative purpose, we increase the shape parameter ϑ_2 from 0.2 to 6 with step size 0.2. Fix the mean ϑ_1 at 10. Gradually increase the sample size n from 10 to 100 with step size 5. For each combination of ϑ_2 and n , we generate 1000 data sets. Calculate $\ell_{\hat{\theta}}^n$, $\ell_{\hat{\vartheta}}^n$ and ζ_{Ga}^n for each data set. For the 1000 pairs of $\{\ell_{\hat{\theta}}^n, \ell_{\hat{\vartheta}}^n\}$, calculate the percentage of $\ell_{\hat{\theta}}^n$ being larger than $\ell_{\hat{\vartheta}}^n$. The percentage is an estimate of τ_n . Take average of the 1000 values of ζ_{Ga}^n . 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 τ_n and the averaged value of ζ_{Ga}^n in Figure 1. In Figure 1, the red curve corre-

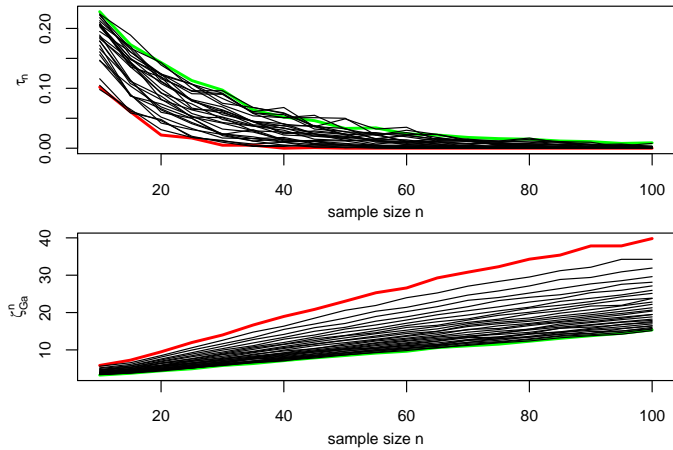


Figure 1: The evolution of τ_n (above) and the averaged value of ζ_{Ga}^n (blow).

sponds to $\vartheta_2 = 0.2$, and the green curve corresponds to $\vartheta_2 = 6$. Figure 1 shows that, if the shape parameter ϑ_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, τ_n should be smaller than 0.05. We let n_1 denote the sample size, beyond which the value of τ_n is smaller than 0.05; let n_2 denote the sample size, beyond which the averaged value of ζ_{Ga}^n is larger than 7.815. The evolution of n_1 and n_2 along with ϑ_2 is summarized in Table 1. Table 1 shows that selecting

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

ϑ_2	$n_1(\tau_n)$	$n_2(\zeta_{Ga}^n)$	ϑ_2	$n_1(\tau_n)$	$n_2(\zeta_{Ga}^n)$
0.2	20	20	3.4	30	35
0.6	20	20	3.8	35	35
1.0	20	25	4.2	35	40
1.4	25	25	4.6	40	40
1.8	25	30	5.0	45	45
2.2	30	30	5.4	45	45
2.6	35	35	5.8	45	45
3.0	30	35	6.0	45	45

model according to maximized log-likelihoods is as efficient as selecting model according to the proposed test statistic.

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 $f_{Ga}(x; \theta)$. Fit the inverse Gaussian distribution to the data X_n . Maximize the quasi log-likelihood to obtain the QML estimate of ϑ , i.e. $\hat{\vartheta}$. The corresponding maximized quasi log-likelihood is denoted by $\ell_{\hat{\vartheta}}^n$. Given the data X_n and the QML estimate $\hat{\vartheta}$, calculate the value of the test statistic ζ_{IG}^n . For comparison, fit the gamma distribution to the data X_n , maximize the quasi log-likelihood, and denote the maximized quasi log-likelihood by $\ell_{\hat{\theta}}^n$. Define a statistic τ_n :

$$\tau_n = P(\ell_{\hat{\vartheta}}^n > \ell_{\hat{\theta}}^n) = P(n\ell_{\hat{\vartheta}}^n > n\ell_{\hat{\theta}}^n).$$

For a given sample size n , τ_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 θ_1 from 1.8 to 7 with step size 0.2. (Experiments showed that, when $\theta_1 < 1.8$ and n is small, the matrix $V_{IG}^n(\hat{\vartheta})$ is likely to be singular. Hence, we start from 1.8 instead of from 0.2.) Fix the scale parameter θ_2 at 1. Gradually increase the sample size n from 10 to 100 with step size 5. For each combination of θ_1 and n , we generate 1000 data sets. Calculate $\ell_{\hat{\vartheta}}^n$, $\ell_{\hat{\theta}}^n$ and ζ_{IG}^n for each data set. For the 1000 pairs of $\{\ell_{\hat{\vartheta}}^n, \ell_{\hat{\theta}}^n\}$, calculate the percentage of $\ell_{\hat{\theta}}^n$ being smaller than $\ell_{\hat{\vartheta}}^n$. The percentage is an estimate of τ_n . Take average of the 1000 values of ζ_{IG}^n .

Plot the evolution of τ_n and the averaged value of ζ_{IG}^n in Figure 2. In Figure 2, the red curve corresponds

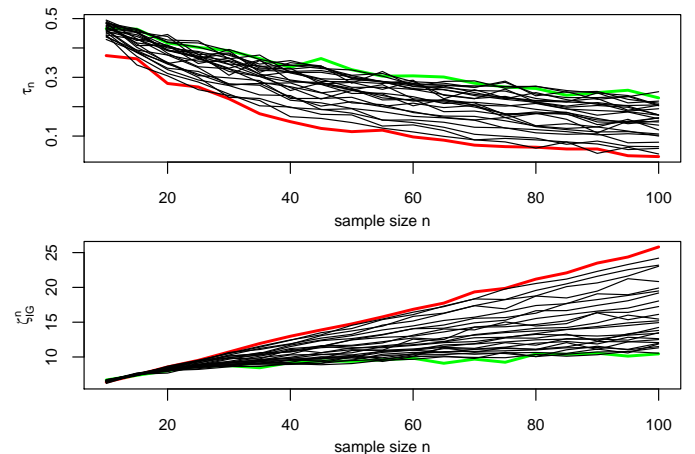


Figure 2: The evolution of τ_n (above) and the averaged value of ζ_{IG}^n (blow).

to $\theta_1 = 1.8$, and the green curve corresponds to $\theta_1 = 7$. The upper panel in Figure 2 shows that, if the shape parameter θ_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 τ_n is smaller than 0.05; let n_2 denote the sample size, beyond which the averaged

value of ζ_{IG}^n is larger than 7.815. The evolution of n_1 and n_2 along with θ_1 is summarized in Table 2. From

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

θ_1	$n_1(\tau_n)$	$n_2(\zeta_{IG}^n)$	θ_1	$n_1(\tau_n)$	$n_2(\zeta_{IG}^n)$
1.8	95	20	4.6	>100	20
2.2	90	20	5.0	>100	20
2.6	>100	20	5.4	>100	20
3.0	>100	20	5.8	>100	20
3.4	>100	20	6.2	>100	20
3.8	>100	20	6.6	>100	20
4.2	>100	20	7.0	>100	20

Table 2, it is clear that the proposed test statistic is much more efficient than τ_n . By comparing Figures 1 and 2 and Tables 1 and 2, it can be found that

- from τ_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 ζ_{Ga}^n or ζ_{IG}^n point of view, when fitting a gamma distribution to inverse-Gaussian distributed data, the required sample size increases with the shape parameter ϑ_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 τ_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

- Al Labadi, L. & M. Zarepour (2013). On asymptotic properties and almost sure approximation of the normalized inverse-gaussian process. *Bayesian Analysis* 8(3), 553–568.
- Dufresne, F., H. U. Gerber, & E. S. W. Shiu (1991). Risk theory and the gamma process. *ASTIN Bulletin* 22, 177–192.
- Griffin, P. S., R. A. Maller, & D. Roberts (2013). Finite time ruin probabilities for tempered stable insurance risk processes. *Insurance: Mathematics and Economics* 53(2), 478–489.
- Hyodo, M., T. Yamada, & M. S. Srivastava (2012). A model selection criterion for discriminant analysis of high-dimensional data with fewer observations. *Journal of Statistical Planning and Inference* 142(12), 3134–3145.
- Kullback, S. & R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics* 22(1), 79–86.
- Tang, Y., W. Pei, K. Wang, Z. He, & Y. Cheung (2009). A theoretical analysis of multiscale entropy under the inverse gaussian distribution. *International Journal of Bifurcation and Chaos* 19(09), 3161–3168.
- Tsai, C., S. Tseng, & N. Balakrishnan (2011). Mis-specification analyses of gamma and wiener degradation processes. *Journal of Statistical Planning and Inference* 141(12), 3725–3735.
- van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety* 94(1), 2–21.
- Wasan, M. T. (1968). On an inverse gaussian process. *Scandinavian Actuarial Journal* 1968(1-2), 69–96.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50(1), 1–25.
- Zhang, M., Z. Ye, & M. Xie (2014). A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering* 77, 103 – 114.
- Zhang, S. & W. Zhou (2014). Cost-based optimal maintenance decisions for corroding natural gas pipelines based on stochastic degradation models. *Engineering Structures* 74, 74 – 85.
- Zhou, Q. M., P. X. Song, & M. E. Thompson (2012). Information ratio test for model misspecification in quasi-likelihood inference. *Journal of the American Statistical Association* 107(497), 205–213.
- Zhu, W., M. Fouladirad, & C. Brenguer (2015). Condition-based maintenance policies for a combined wear and shock deterioration model with covariates. *Computers & Industrial Engineering* 85, 268 – 283.