

# 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. Hence, one challenge for practitioners is determining which of the two processes is most appropriate in light of a real data set. 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. We conduct extensive numerical study to demonstrate the efficiency (in terms of sample size) of the proposed test statistic. We also indicate the conditions under which a gamma process can be well approximated by an 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. Hence, both the gamma process and the inverse Gaussian process are appropriate for stochastic modelling of monotonic and gradual deterioration. The gamma processes were satisfactorily fitted to data on creep of concrete, fatigue crack growth, corroded steel gates, thinning due to corrosion, and chloride ingress into concrete; see van Noortwijk (2009) for a review. Applications of the inverse Gaussian process can be found in Al Labadi and Zarepour (2013), Ivanov (2013), Zhang et al. (2013), Zhang et al. (2014) and Peng (2015).

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:

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

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

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

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  $u^3/\lambda$ . 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. 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 Bachoc (2013), Richmond and Horowitz (2015) and Spokoiny and Zhilova (2015). This paper proposes an efficient and broadly applicable test statistic for model selection by quoting the theorems in White (1982). The test statistic is built on the notion of information equivalence: under mild regularity conditions, the information matrix can be expressed in either the Hessian form or the outer product form. Hence, the difference between the two information matrices will indicate whether or not the model is misspecified.

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 similar.

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,  $Ga(\alpha\Delta, \beta)$ . 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  $\theta$ :

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

Find a value of  $\theta$  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 maximization of  $\ell_{Ga}(\theta; X_n)$  is analogous to the minimization of the Kullback-Leibler divergence of  $F_{Ga}(x; \theta)$  from  $G(x)$ . 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 very 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  $\theta$ . 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  $\theta^*$  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  $\theta^*$  is inaccessible. It is clear that  $\hat{\theta}$  is a consistent estimator for  $\theta^*$ , because  $\ell_{G_a}(\theta; X_n)$  is a consistent estimator for  $E[\log(f_{G_a}(x; \theta))]$ .

Define two  $2 \times 2$  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  $\theta^0$ ;
- $\sqrt{n}(\hat{\theta} - \theta^0)$  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  $3 \times 2$  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  $\theta$  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 “ $\nabla$ ” denotes the vector differential operator. Define two  $3 \times 3$  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  $\chi_3^2$  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.

**Remark 1** *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

To demonstrate the efficiency of the proposed test statistic, we study the problem of model selection between the gamma process and the inverse Gaussian process. In Section 3.1, we simulate data from the inverse Gaussian process but fit the gamma process to the simulated data. In Section 3.2, we simulate data from the gamma process but fit the inverse Gaussian process to the simulated data.

#### 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  $\theta$ , i.e.  $\hat{\theta}$ . Given the data  $X_n$  and the QML estimate  $\hat{\theta}$ , calculate the value of the test statistic  $\zeta_{G_a}^n$ . The critical value of the chi-square distribution with 3 degrees of freedom at significance level 0.05 is 7.815. Hence, if  $\zeta_{G_a}^n$

is larger than 7.815, we say that the gamma process is inappropriate for the data set.

The parameter setting is described as follows. We increase the shape parameter  $\vartheta_2$  from 0.2 to 6 with step size 0.2. The mean  $\vartheta_1$  takes in turn the following three values:  $\{1, 10, 50\}$ . (Experiments showed that, when  $\vartheta_1 > 60$  and  $n$  is small, the matrix  $V_{Ga}^n(\hat{\theta})$  is likely to be singular.) Gradually increase the sample size  $n$  from 20 to 100 with step size 5. For each combination of  $\vartheta_1$ ,  $\vartheta_2$  and  $n$ , we generate 1000 data sets. Calculate  $\zeta_{Ga}^n$  for each data set. Two statistics are studied: (1) the average of the 1000 values of  $\zeta_{Ga}^n$ , and (2) the percentage of  $\zeta_{Ga}^n$  being larger than 7.815.

Plot the evolution of the two statistics in Figure 1. In Figure 1, the red curve (with circular points) corresponds to  $\vartheta_2 = 0.2$ , and the green curve (with square points) corresponds to  $\vartheta_2 = 6$ . The dashed line in the left three panels corresponds to the critical value 7.815, and in the right three panels corresponds to percentage 0.5. The top two panels correspond to  $\vartheta_1 = 1$ ; the middle two panels correspond to  $\vartheta_1 = 10$ ; the bottom two panels correspond to  $\vartheta_1 = 50$ .

The green curves in the top two panels show that, even if the sample size is 100, there is only a small percentage of the 1000  $\zeta_{Ga}^n$ 's whose values are larger than 7.815. The red curves in the bottom two panels show that, even if the sample size is only 20, there is a large percentage of the 1000  $\zeta_{Ga}^n$ 's whose values are larger than 7.815. Hence, we can conclude that an inverse Gaussian process, with a large shape parameter and a small mean, can be closely approximated by a gamma process. In Figure 1, when the mean value is relatively large, the percentage  $\Pr(\zeta_{Ga}^n > 7.815)$  increases rapidly with the sample size, implying that the test statistic is efficient.

Table 1 lists, for each combination of  $\vartheta_1$  and  $\vartheta_2$ , the

Table 1: The required sample size for the expected value of  $\zeta_{Ga}^n$  being larger than 7.815.

$\vartheta_2$	$\vartheta_1$			$\vartheta_2$	$\vartheta_1$		
	1	10	50		1	10	50
0.2	30	20	20	3.4	>100	35	20
0.6	50	20	20	3.8	>100	35	25
1.0	65	25	20	4.2	>100	40	25
1.4	75	25	20	4.6	>100	40	25
1.8	95	30	20	5.0	>100	45	25
2.2	>100	30	20	5.4	>100	45	25
2.6	>100	35	20	5.8	>100	45	25
3.0	>100	35	20	6.0	>100	50	25

required sample size for the expected value of  $\zeta_{Ga}^n$  being larger than 7.815. Table 1 further verifies that an inverse Gaussian process, with a large shape parameter and a small mean, can be closely approximated by a gamma process. In Table 1, when the shape parameter is small and the mean is relative large (e.g.,  $\vartheta_2 = 1.4$  and  $\vartheta_1 = 10$ ), the required sample size for the expected value of  $\zeta_{Ga}^n$  being larger than 7.815 is very small. When  $\vartheta_1 = 50$ , for all the studied values

of  $\vartheta_2$ , the required sample size is always small, showing the efficiency of the proposed test statistic in such case.

### 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  $\vartheta$ , i.e.  $\hat{\vartheta}$ . Given the data  $X_n$  and the QML estimate  $\hat{\vartheta}$ , calculate the value of the test statistic  $\zeta_{IG}^n$ .

We increase the shape parameter  $\theta_1$  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.) The scale parameter  $\theta_2$  takes in turn the following three values:  $\{1, 10, 100\}$ . Gradually increase the sample size  $n$  from 20 to 100 with step size 5. For each combination of  $\theta_1$ ,  $\theta_2$  and  $n$ , we generate 1000 data sets. Calculate  $\zeta_{IG}^n$  for each data set. Likewise, two statistics are studied: (1) the average of the 1000 values of  $\zeta_{IG}^n$ , and (2) the percentage of  $\zeta_{IG}^n$  being larger than 7.815.

Plot the evolution of the two statistics in Figure 2. In Figure 2, the red curve (with circular points) corresponds to  $\theta_1 = 1.8$ , and the green curve (with square points) corresponds to  $\theta_1 = 7$ . The dashed line in the left three panels corresponds to the critical value 7.815, and in the right three panels corresponds to percentage 0.5. The top two panels correspond to  $\theta_2 = 1$ ; the middle two panels correspond to  $\theta_2 = 10$ ; the bottom two panels correspond to  $\theta_2 = 100$ .

Form Figure 2 it is observed that  $\zeta_{IG}^n$  is insensitive to the value of the scale parameter  $\theta_2$ : the three left panels are quite similar and so are the right three panels. In Figure 2, for some large values of  $\theta_1$ , the percentage  $\Pr(\zeta_{IG}^n > 7.815)$  increases slowly with the sample size. But most of the percentage curves are above the 0.5 line, implying that we are prone to reject the misspecified model. The green curves in Figure 2 further show that a gamma process with a large shape parameter can be closely approximated by an inverse Gaussian process.

Table 2 lists, for each combination of  $\theta_1$  and  $\theta_2$ , the

Table 2: The required sample size for the expected value of  $\zeta_{IG}^n$  being larger than 7.815.

$\theta_1$	$\theta_2$			$\theta_1$	$\theta_2$		
	1	10	100		1	10	100
1.8	20	20	20	4.6	20	20	20
2.2	20	20	20	5.0	20	20	20
2.6	20	20	20	5.4	25	20	20
3.0	20	20	20	5.8	25	20	20
3.4	20	20	20	6.2	25	20	20
3.8	20	20	20	6.6	25	20	20
4.2	20	20	20	7.0	25	20	20

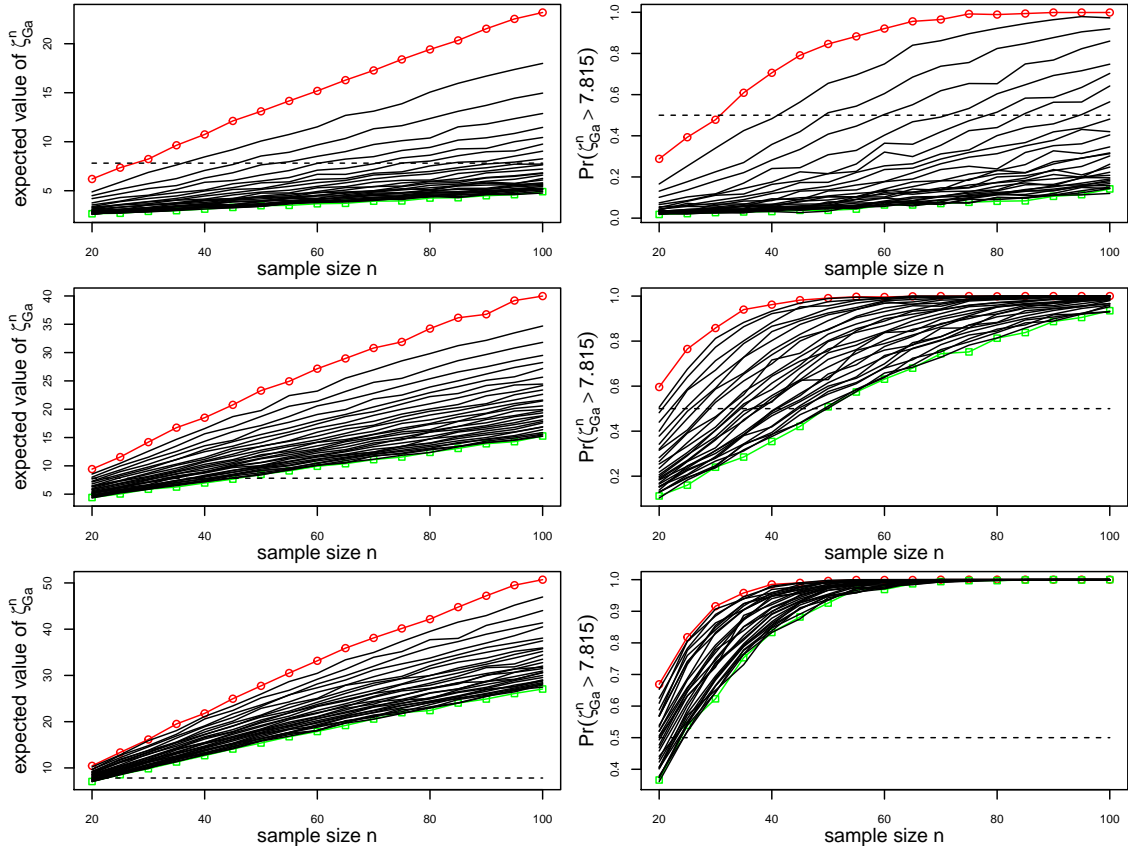


Figure 1: The evolution of the averaged value of  $\zeta_{Ga}^n$ , and the evolution of the probability  $\Pr(\zeta_{Ga}^n > 7.815)$ . In each panel, each curve corresponds to a value of  $\vartheta_2$  (increasing from 0.2 to 6). The top two panels correspond to  $\vartheta_1 = 1$ ; the middle two panels correspond to  $\vartheta_1 = 10$ ; the bottom two panels correspond to  $\vartheta_1 = 50$ . The dashed line in the left three panels corresponds to the critical value 7.815, and in the right three panels corresponds to percentage 0.5.

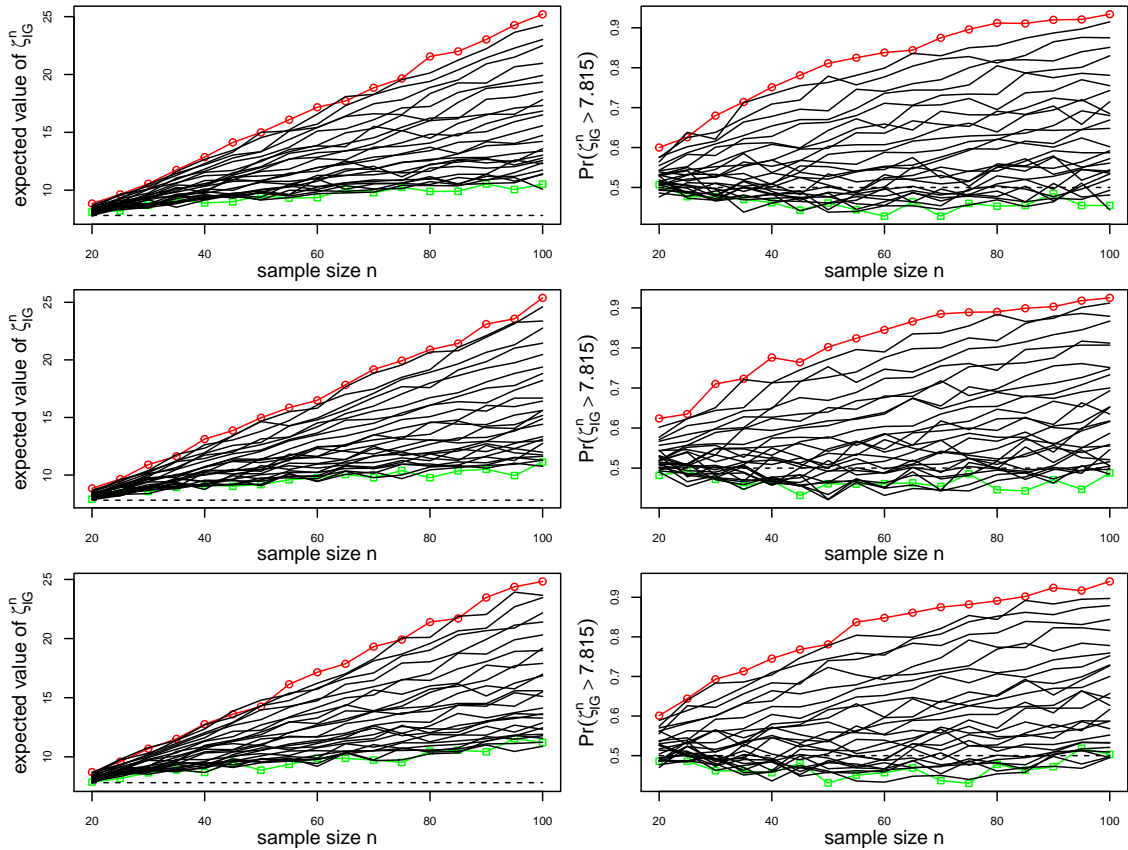


Figure 2: The evolution of the averaged value of  $\zeta_{IG}^n$ , and the evolution of the probability  $\Pr(\zeta_{IG}^n > 7.815)$ . In each panel, each curve corresponds to a value of  $\theta_1$  (increasing from 1.8 to 7). The top two panels correspond to  $\theta_2 = 1$ ; the middle two panels correspond to  $\theta_2 = 10$ ; the bottom two panels correspond to  $\theta_2 = 100$ . The dashed line in the left three panels corresponds to the critical value 7.815, and in the right three panels corresponds to percentage 0.5.

required sample size for the expected value of  $\zeta_{IG}^n$  being larger than 7.815. From the expected value point of view, the proposed test statistic is quite efficient for testing if an inverse Gaussian process is appropriate for a given data set, and Table 2 implies that the inverse Gaussian distribution is not capable of well fitting a gamma distributed data.

#### 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. It was found that an inverse Gaussian process with a large shape parameter and a small mean can be well approximated by a gamma process. A gamma process with a large shape parameter can also be closely approximated by an inverse Gaussian process. In general, the gamma process is more flexible than the inverse Gaussian process. More simulation studies need be conducted to compare the proposed test statistic with other test statistics.

#### 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.
- Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. *Computational Statistics & Data Analysis* 66, 55 – 69.
- Dufresne, F., H. U. Gerber, & E. S. W. Shiu (1991). Risk theory and the gamma process. *ASTIN Bulletin* 22, 177–192.
- Ivanov, R. V. (2013). Closed form pricing of european options for a family of normal-inverse gaussian processes. *Stochastic Models* 29, 435 – 450.
- Kullback, S. & R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics* 22(1), 79–86.
- Peng, C.-Y. (2015). Inverse gaussian processes with random effects and explanatory variables for degradation data. *Technometrics* 57(1), 100–111.
- Richmond, C. D. & L. L. Horowitz (2015). Parameter bounds on estimation accuracy under model misspecification. *IEEE Transactions on Signal Processing* 63(9), 2263–2278.
- Spokoiny, V. & M. Zhilova (2015). Bootstrap confidence sets under model misspecification. *The Annals of Statistics* 43(6), 2653–2675.
- 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, & H. Qin (2013). Inverse gaussian process-based corrosion growth model for energy pipelines considering the sizing error in inspection data. *Corrosion Science* 73, 309 – 320.