On Prequential omnibus Goodness-of-fit Tests for
discrete-time Stochastic Processes

Mhamed-Ali El-Aroui
ISG de Tunis
41 rue de la liberté
Bardo 2000, Tunisia
Mhamed.Elaroui@isg.rnu.tn

Olivier Gaudoin
INP Grenoble
Laboratoire IMAG-LMC
38041 Grenoble Cedex 9, France
Olivier.Gaudoin@imag.fr

Abstract

The present paper is a contribution to the study of prequential Goodness-of-fit (Gof) tests for discrete-time stochastic processes. The study focuses on an omnibus Gof test based on Rosenblatt Probability Integral Transform (RPIT) within Dawid’s prequential framework. This Gof test is easy to use and the test statistics has the same asymptotic distribution under a wide range of null hypotheses. It is shown that other prequential approaches for model fitting can be related to RPIT. Intensive Monte-Carlo simulations are presented to investigate the behavior of this test for a wide range of stochastic process models: renewal processes, autoregressive processes, non homogeneous Poisson processes, generalized linear models... These simulations strongly suggest that the RPIT test could be used to test the fit of these models. It is also conjectured that this test is still appropriate for testing the Gof of any discrete-time stochastic process provided that efficient estimators are used. Finally, the power of the RPIT Gof test is studied and happens to be satisfactory.

Key words: Model Choice, Rosenblatt Probability Integral Transform, Predictive performance, Plug-in Statistical Forecasting System, Prequential Efficiency and Consistency.

AMS subject classification – 62L12, 62F03, 62F05.

1 Introduction

Let \( X \equiv \{ X_i \}_{i \in \mathbb{N}} \) be a sequence of random variables defined on a filtered probability space \( (\Omega, \mathcal{F}, \{ \mathcal{F}_i \}_{i \geq 1}, \mathbb{P}) \). At each step \( i \) a forecaster is asked to predict the next observation \( X_{i+1} \) using the available information \( \mathcal{F}_i \). This paper considers the problem of selecting statistical models for \( X \) with regard to the quality of their forecasts. A statistical model may be defined by a family of parametric distributions \( \mathcal{P} = \{ P_\theta ; \theta \in \mathcal{T} \subset \mathbb{R}^k \} \). The main interest here is to find goodness-of-fit (Gof) tests for \( \mathcal{P} \), i.e. statistical tests for the null hypothesis \( H_0 : \mathbb{P} \in \mathcal{P} \) vs \( H_1 : \mathbb{P} \notin \mathcal{P} \) where test statistics have the same asymptotic behavior for a wide range of model families. This will provide a general and rigorous statistical tool for comparing forecasts from competing models.

Several criteria for model selection are available in the literature: AIC (Akaike, 1974), Cp
Dawid predictive-sequential (prequential) approach (see Dawid, 1984) is used in this work: observations are supposed to arrive sequentially and outcomes $x^{(i)} = (x_1, \ldots, x_i)$ are used to predict the next observation $X_{i+1}$. Dawid terminology is also used: a probability forecasting system (PFS) is any rule that, for any $i \geq 0$ and any set of outcomes $x^{(i)}$, gives a predictive probability distribution $P^*_{i+1}$ for $X_{i+1}$. When the statistical model $\mathcal{P}$ is used to generate $P^*_{i+1}$, the PFS is called statistical forecasting system (SFS). Any given PFS (or SFS) is consistent with one and only one probability measure $P^*$ on $\Omega$. A point prediction system (PPS) is any rule that, for any $i \geq 0$ and any set of outcomes $x^{(i)}$, gives a point forecast $x^*_{i+1}$ for $X_{i+1}$. Let $P_{\theta, X_{i+1}}(\cdot | X^{(i)} = x^{(i)})$ denote the conditional distribution of $X_{i+1}$ given $X^{(i)} = x^{(i)}$ under $P_{\theta}$.

Examples of SFS are:

— Plug-in SFS: the predictive distribution at step $i$ (denoted $P^*_{i+1}$) is obtained by replacing the unknown parameter $\theta$ by an estimation $\hat{\theta}_i$ based on the available outcomes $x^{(i)}$:

$$P^*_{i+1}(\cdot) = P_{\hat{\theta}_i, X_{i+1}}(\cdot | X^{(i)} = x^{(i)}).$$

— Bayesian SFS: a prior distribution $\pi$ is assigned to $\theta$, the predictive distribution for $X_{i+1}$ is:

$$P^*_{i+1}(\cdot) = \int_{\theta \in \mathcal{T}} P_{\theta, X_{i+1}}(\cdot | X^{(i)} = x^{(i)}) \pi(\theta | X^{(i)} = x^{(i)}) \, d\theta.$$  

In all what follows, $P^*$ will denote a particular SFS issued from $\mathcal{P}$. The main question of this work is how to use predictive distributions $P^*_{i+1}$ (or point forecasts $x^*_{i+1}$) and the future outcomes $x_{i+1}$ to assess the statistical validity of the parametric model $\mathcal{P}$. Skouras and Dawid (1998) used the sum of squared prediction errors $S(P, n) = \sum_{i=1}^{n} (x_i - x^*_i)^2$ to define the efficiency of a PFS $P$ for the model $\mathcal{P}$: $P$ is efficient for $\mathcal{P}$ if for any other PFS $Q$, $\lim_{n \to +\infty} \sup_{P(Q, n)} (S(P, n) - S(Q, n)) < +\infty$ with $P_{\theta}$-probability 1 for almost all $\theta \in \mathcal{T}$. Nevertheless, this notion of efficiency does not provide tools for choosing between two efficient models. It does neither provide statistical tests for the hypothesis $H_0 : \{P \in \mathcal{P}\}$.

Seillier-Moiseiwitsch et al. (1992) suggested prequential Gof tests based on a martingale central limit theorem (MCLT). The asymptotic normality of their test statistics holds under technical (Lindeberg-type) conditions.

Arjas and Gasbarra (1997) presented a prequential procedure to assess the predictive performance of SFS for marked point processes. They used a result proving the exponential distribution and independence of the compensator spacings under $\mathbb{P} = P^*$ (but not under $\mathbb{P} = P_{\theta} \in \mathcal{P}$).
A third approach of prequential model assessment uses Rosenblatt probability integral transform (RPIT) (see Rosenblatt, 1952 and Dawid, 1984). This transform gives quantities \( u_1, \ldots, u_n \), which form, approximately, a sample from Uniform\([0,1]\) distribution if \( P^* \) is “close” to \( \mathbb{P} \). This gives an empirical tool to assess the quality of SFS based on \( \mathcal{P} \) but still does not give rigorous statistical Gof tests for \( H_0 : \mathbb{P} \in \mathcal{P}^* \).

From Monte-Carlo simulation studies in the case of reliability models (Downs and Scott, 1992 and Crétois et al., 1999) it seems that the RPIT approach can lead to formal prequential Gof tests. This was formally proven for the homogeneous Poisson process (HPP) (El-Aroui, 1999) and for a particular non homogeneous Poisson process, namely the Power-Law process (see Crétois et al., 1999). This paper continues the study of RPIT Gof test when applied to test other stochastic process models: renewal, Poisson, autoregressive processes and generalized linear models.

Section 2 presents Dawid’s prequential framework and two prequential methods for comparing predictive quality of competing models. The proposed omnibus Gof test based on Rosenblatt transform is described in section 3. Links between RPIT and other prequential model fitting approaches are studied in section 4. It is conjectured that the RPIT Gof test is still appropriate for testing the Gof of any model provided that efficient estimators are used. In section 5, statistical properties of the RPIT Gof test are numerically studied for renewal, Poisson, autoregressive processes and generalized linear models. The power of the RPIT Gof test is discussed. Finally, conclusions and open questions are presented in section 6.

2 Testing Model-fit within the prequential approach

Many empirical tools were proposed by Dawid and co-authors (Dawid, 1984, Seillier-Moiseiwitsch et al., 1992 and 1993, Skouras and Dawid, 1998 and 1999) to compare predictive performance of competing models. It will be shown in the following that the RPIT approach which was presented as an empirical tool can lead to formal Gof tests. In this section, we briefly present the prequential Gof approaches proposed by Seillier-Moiseiwitsch et al. (1992 and 1993) and Arjas and Gasbarra (1997). Before that, two important notions are presented hereafter: prequential efficiency and consistency.

All the prequential tools for model selection use either “prequentially efficient” or “prequentially consistent” SFS’s. The notion of prequential efficiency (first defined in Dawid, 1984, and deeply studied in Skouras and Dawid, 1999) may be summarized by saying that a SFS \( P \) is \textit{efficient} with respect to \( \mathcal{P} \) if there is no other SFS giving “better” forecasts than \( P \) almost
surely for all $\theta \in \mathcal{T}$. This means that $P$ cannot be beaten as long as data are generated by some $P_0$ in $\mathcal{P}$. A SFS $P$ is sequentially consistent for $\mathcal{P}$ if a suitable measure of the distance between the one step ahead predictive distributions $P_i$ and $P_i^0$ converges to 0 under $P_0$, for almost all $\theta$. Skouras and David (1999) conjectured that whenever there exists a consistent SFS, then any efficient one will be consistent, and the converse is false.

2.1 Prequential Gof approaches based on martingale central limit theorem (MCLT)

Seillier-Moiseiwitsch et al. (1992 and 1993) proposed prequential Gof tests for which test statistics, whose asymptotic normality is investigated, have the following form:

$$Y_n = \frac{\sum_{i=1}^{n} (S_i - \mu_i)}{(\sum_{i=1}^{n} \sigma_i^2)^{1/2}}$$

(1)

where for any non-negative integer $i$, $S_i$ is a function of $X^{(i)}$, $\mu_i = \mathbb{E}_{P^*}(S_i|X^{(i-1)})$ and $\sigma_i^2 = \text{Var}_{P^*}(S_i|X^{(i-1)})$.

Since under $P^*$ the sequence $\left\{ \sum_{i=1}^{n} (S_i - \mu_i) \right\}_{n \geq 1}$ forms a martingale adapted to $\mathcal{F}_n$, Seillier-Moiseiwitsch et al. used a standard martingale central limit theorem to prove that, under Lindeberg-type conditions, $Y_n$ converges in distribution, as $n \to \infty$, to the standard normal distribution:

$$Y_n \xrightarrow{D,n \to \infty} \mathcal{N}(0,1) \quad \text{under } P^*.$$

(2)

Nevertheless, the previous result will allow checking \textquoteleft\textquoteleft $\mathbb{P} = P^*$\textquoteright\textquoteright but not testing the hypothesis $H_0 : \mathbb{P} \in \mathcal{P}$ for which one needs to prove:

$$Y_n \xrightarrow{D,n \to \infty} \mathcal{N}(0,1) \quad \text{under } P_0 \quad \text{for almost all } P_0 \in \mathcal{P}.$$  

(3)

When $P^*$ is a Bayesian SFS, Seillier-Moiseiwitsch et al. (1992) used a heuristic argument, presented hereafter, to suggest that (3) \textquoteleft\textquoteleft will hold, under suitable smoothness conditions, in very great generality\textquoteright\textquoteright. Let $\pi$ denote a prior density on $\mathcal{T}$ and $\Theta$ denote the r.v. modeling the uncertainty on parameter $\theta$ within the Bayesian SFS. The heuristic argument used by Seillier-Moiseiwitsch et al. is based on the following idea: the martingale limit theorem used to prove (2) is mixing, that is for any fixed event $E$ and all real $x$, $P^*[(Y_n \leq x) \cap E] \xrightarrow{n \to \infty} \Phi(x) P(E)$ where $\Phi$ is the Cumulative distribution function (Cdf) of the standard Normal distribution. Thus $Y_n$ is asymptotically independent of any fixed event. If $N(\theta)$ is an open neighborhood of $\theta$ and $E \equiv \Theta \in N(\theta)$, then the mixing property implies $P^*[Y_n \leq x | \Theta \in N(\theta)] \xrightarrow{n \to \infty} \Phi(x)$. As $N(\theta)$ shrinks to $\theta$, the left hand-side approaches $P_0(Y_n \leq x)$. Seillier-Moiseiwitsch et al. concluded
that under conditions justifying the interchange of the two limiting operations, one could deduce (3).

Seillier-Moiseiwitsch et al. (1992) showed that the former heuristic argument still holds for efficient non-Bayesian SFS $\mathcal{P}$. This is due to the fact that for any efficient SFS $P^*$, there exists a Bayesian SFS $P_B$ which is absolutely continuous with respect to $P^*$ Then, a Rényi result proving that a mixing limit theorem is unaffected by an absolutely continuous change of measure is used to deduce that $Y_n \xrightarrow{D,n \to \infty} N(0,1)$ under $P_B$. The heuristic argument can be used again since a Bayesian SFS is considered.

The previous prequential tests are very promising since, up to the checking of Lindeberg conditions, they probably would give omnibus Gof tests for a wide range of stochastic models.

### 2.2 Prequential Gof approach based on compensator spacings

Arjas and Gasbarra (1997) gave empirical tools for assessing forecasts from life history models. They considered a point process defined by its times between events $X \equiv \{X_i\}_{i \in \mathbb{N}}$. Let $0 = T_0 < T_1 < \ldots$ denote times of event occurrences and $\{N_t\}_{t \geq 0}$ denote the related counting process. For a time $t \geq 0$, let: $H_t \equiv \{T_i : T_i \leq t\}$, $H_{t-} \equiv \{T_i : T_i < t\}$ and $\mathcal{H}_t$ denote the internal filtration of the point process: $\mathcal{H}_t = \sigma\{H_t\}$. For a SFS $P^*$, Arjas and Gasbarra used the predictive $(P^*, \mathcal{H}_t)$-intensity $\hat{\lambda}_t$ defined by $\hat{\lambda}_t = \lim_{dt \to 0} \frac{1}{dt} P^*\{N_{t+dt} - N_t = 1|H_{t-}\}$. The related $(P^*, \mathcal{H}_t)$-compensators are denoted $\hat{\Lambda}_t = \int_0^t \hat{\lambda}_s \, ds$.

To assess the predictive quality of the considered models, Arjas and Gasbarra suggested the use of Norros (1986) result on the compensator spacings. This result states that, under $P^*$, the spacings between $(P^*, \mathcal{H}_t)$-compensators: $[\hat{\Lambda}_{T_k} - \hat{\Lambda}_{T_{k-1}}]_{k \geq 1}$ are independent and have the same exponential distribution with parameter 1. The assessment of the predictive performance of $\mathcal{P}$ is reduced to checking the fit of $\hat{\Lambda}_{T_k} - \hat{\Lambda}_{T_{k-1}}$ to this exponential distribution.

Prequential model fitting proposed by Arjas and Gasbarra checks if "$\mathbb{P} \preceq P^*$" but does not give rigorous statistical Gof tests for $H_0 : \"\mathbb{P} \in \mathcal{P}\"$.

### 3 Prequential Gof approach based on Rosenblatt Probability Integral Transform (RPIT)

We present in this section the conditional probability integral transform (RPIT) of Rosenblatt (1952) and its use for the definition of an omnibus Gof test.
3.1 Rosenblatt Probability Integral Transform

Let us assume that the distribution of $X$ is absolutely continuous and let $R_{\mathbb{P}}$ be the transformation defined by $R_{\mathbb{P}}(x_1,\ldots,x_n) = (u_1,\ldots,u_n)$ where the $u_i$'s are given for $(x_1,\ldots,x_n) \in \mathbb{R}^n$ by:

$$u_1 = \mathbb{P}(X_1 \leq x_1), \quad u_2 = \mathbb{P}(X_2 \leq x_2 | X_1 = x_1); \quad \ldots \quad u_n = \mathbb{P}(X_n \leq x_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}).$$

The $u_i$'s may be considered as realizations of random variables $U_i$'s given by $(U_1,\ldots,U_n) = R_{\mathbb{P}}(X_1,\ldots,X_n)$. If we consider the following conditional Cdf's:

$$F_1(x_1) = \mathbb{P}(X_1 \leq x_1); \quad F_2|_{x_2}(x_2) = \mathbb{P}(X_2 \leq x_2 | X_1 = x_1); \quad \ldots$$

$$F_{n|_{x_1,\ldots,x_{n-1}}}(x_n) = \mathbb{P}(X_n \leq x_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1})$$

then the r.v.'s $U_i$ can be defined by

$$U_1 = F_1(X_1); \quad U_2 = F_2|_{x_2}(X_2); \quad \ldots \quad U_n = F_{n|_{x_1,\ldots,x_{n-1}}}(X_n).$$

Rosenblatt (1952) showed that, under $\mathbb{P}$, $U_i$'s are i.i.d and uniformly distributed on [0,1]. Dawid (1984) suggested to assess the predictive quality of a SFS $P^*$ by computing $(\tilde{u}_1,\ldots,\tilde{u}_n) = R_{P^*}(x_1,\ldots,x_n)$ and examining if $\tilde{u}_i$'s “look like” a random sample from a Uniform[0,1]. This would suggest that $P^*$ is close to $\mathbb{P}$. Littlewood and co-authors (Keiller et al., 1983, and Abdel-Ghali et al., 1986) used RPIT to compare software reliability models. The plot giving the empirical distribution function of the $\tilde{u}_i$'s is called U-plot. As mentioned by Littlewood and co-authors, the RPIT approach gives an empirical criterion for comparing models but not a formal statistical Gof test. This is due to the fact that, under $H_0 : \mathbb{P} = P_\theta \in \mathcal{P}$, outcomes $u_i$'s of $(U_1,\ldots,U_n) = R_{P_\theta}(X_1,\ldots,X_n)$ (which are i.i.d Uniform[0,1]) cannot be computed since $\theta$ is unknown. Moreover, there is no apparent reason for $(\tilde{U}_1,\ldots,\tilde{U}_n) = R_{P^*}(X_1,\ldots,X_n)$ nor $(\tilde{U}_1,\ldots,\tilde{U}_n) = R_{P_{\hat{\theta}_n}}(X_1,\ldots,X_n)$ to be i.i.d Uniform[0,1] under $\mathbb{P} = P_\theta$ (this is true respectively under $\mathbb{P} = P^*$ and $\mathbb{P} = P_{\hat{\theta}_n}$ but not under $\mathbb{P} = P_\theta$).

3.2 RPIT prequential Gof test for plug-in SFS

When $P^*$ is a plug-in SFS, the realization of $\tilde{U}_i$ is

$$\tilde{u}_i = P^*(X_i \leq x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = P_{\hat{\theta}_{i-1}}(X_i \leq x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$$

where $\hat{\theta}_{i-1} = \hat{\theta}(x_1,\ldots,x_{i-1})$ is an appropriate estimator of $\theta$. Even if $\tilde{U}_i$'s are i.i.d Uniform[0,1] under $\mathbb{P} = P^*$ but not under $\mathbb{P} = P_\theta$, Dawid and co-workers suggested that if the SFS $P^*$ is prequentially efficient, we will have under $\mathbb{P} = P_\theta$: $P^* \simeq P_\theta$, and $\tilde{U}_i$'s will be “approximately” i.i.d Uniform[0,1].
Indeed, what is needed for formally testing the Gof of the parametric family \( \mathcal{P} \) is the asymptotic behavior of a test statistic based on the \( \tilde{U}_i \)'s. We study here the use of Empirical distribution function (Edf) Gof tests, i.e. tests based on a distance between the Empirical distribution function of the \( \tilde{U}_i \)'s and the Cdf of Uniform[0,1] distribution. Let \( y_n \) and \( \tilde{y}_n \) be respectively the empirical process of the \( U_i \)'s and the \( \tilde{U}_i \)'s:

\[
\forall t \in [0,1] \quad y_n(t) = \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^{n} 1(\tilde{U}_i \leq t) - t \right] \quad \text{and} \quad \tilde{y}_n(t) = \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^{n} 1(\tilde{U}_i \leq t) - t \right]
\]  

(4)

It is well known that

\[
y_n \xrightarrow{D,n \rightarrow \infty} \mathbb{B} \quad \text{under } P_0 \quad \text{and} \quad \tilde{y}_n \xrightarrow{D,n \rightarrow \infty} \mathbb{B} \quad \text{under } P^*.
\]

(5)

where \( \mathbb{B} \) denotes the brownian bridge. The RPIT approach will provide a formal Gof test for the parametric family \( \mathcal{P} \) if it is proven that:

\[
\tilde{y}_n \xrightarrow{D,n \rightarrow \infty} \mathbb{B} \quad \text{under } P_0.
\]

(6)

Let \( \tilde{K}_n \) be the Kolmogorov-Smirnov distance between \( \tilde{U}_i \)'s Edf and the Cdf of the Uniform[0,1] distribution: \( \tilde{K}_n = \sup_{t \in [0,1]} |\tilde{y}_n(t)| \). One could conclude from (6) that

\[
\tilde{K}_n \xrightarrow{D,n \rightarrow \infty} \mathcal{L}_{ks} \quad \text{under } P_0
\]

(7)

where \( \mathcal{L}_{ks} \) is the Kolmogorov-Smirnov probability distribution. This would give a formal Gof test for \( \mathcal{P} \) which consists in computing Kolmogorov-Smirnov distance \( \tilde{K}_n \) (Anderson-Darling or Cramér-von Mises statistics can also be used) and compare it to upper quantiles of \( \mathcal{L}_{ks} \).

When using the previous prequential Gof test in practice, higher efficiency will be obtained if the first \( \tilde{U}_i \)'s (where \( \theta \) is estimated with very few observations \( x_i \)) are not used in the computation of \( \tilde{K}_n \). Therefore \( \tilde{K}_n \) will be replaced by:

\[
\tilde{K}_{n,p} = \sup_{t \in [0,1]} |\tilde{y}_{n,p}(t)| \quad \text{where} \quad \forall t \in [0,1], \quad \tilde{y}_{n,p}(t) = \sqrt{n-p} \left[ \frac{1}{n-p} \sum_{i=p+1}^{n} 1(\tilde{U}_i \leq t) - t \right]
\]

(8)

where \( p \) is a suitably chosen integer. Hence, the key result to be proven becomes:

\[
\tilde{y}_{n,p} \xrightarrow{D,n \rightarrow \infty} \mathbb{B} \quad \text{under } P_0 \quad \text{which implies} \quad \tilde{K}_{n,p} \xrightarrow{D,n \rightarrow \infty} \mathcal{L}_{ks} \quad \text{under } P_0.
\]

(9)

More theoretical work is needed to choose optimal values of \( p \). In the following, we will use \( p = n/2 \).

Several tracks could be tried to prove (9):
- Study the asymptotic properties of \( \hat{U}_i \)'s under \( P_0 \) and deduce \( \hat{y}_{n,p} \xrightarrow{D,n\to\infty} \mathbf{B} \). This approach, used in El-Aroui (1999) to prove (9) for Homogeneous Poisson processes and in Crétois et al. (1999) for a family of Non homogeneous Poisson processes (NHPP), needs specific computations for each tested parametric family.

- Try to formalize Seillier-Moiseiwitsch et al. (1992) heuristic by considering a Bayesian SFS absolutely continuous with respect to the Maximum Likelihood (ML) plug-in SFS \( P^* \) and studying the mixing properties of the standard result: \( y_n \xrightarrow{D,n\to\infty} \mathbf{B} \) under \( P_0 \).

- Prove the finite dimensional convergence of \( \hat{y}_n \) to \( \mathbf{B} \) under \( P_0 \), then prove the tightness of \( \hat{y}_n \).

Section 5 presents extended Monte Carlo simulations which strongly suggest that (9) holds for a wide range of models. This suggests that RPIT approach provides omnibus prequential Gof tests for these models.

4 Linking RPIT to other prequential approaches

4.1 Linking RPIT to the MCLT approach

Prequential model assessment approaches based on MCLT and on RPIT have similar conceptual foundations. In both cases convergence results are available under the SFS probability measure \( P^* \) (see results (2) and (5)). Getting formal prequential Gof tests needs, in both cases, to prove these convergence results under the null hypothesis \( H_0 : "P \in \mathcal{P}" \) when \( P^* \) is a prequentially consistent SFS for \( \mathcal{P} \). Proving the following general conjecture would lead to formal prequential Gof tests issued either from MCLT or RPIT approaches.

**Conjecture** – Let

- \( X \equiv \{X_i\}_{i \in \mathbb{N}} \) be a sequence of random variables defined on a filtered space \( (\Omega, \mathcal{F}, \{\mathcal{F}_i\}_{i \geq 1}, \mathbb{P}) \),

- \( P_1 \) and \( P_2 \) be two probability measures on the previous space

- \( S_i \), for \( i \in \mathbb{N} \), be a scalar function of \( X^{(i)} \),

- \( \mathcal{B} \) be the Borel \( \sigma \)-field on \( \mathbb{R} \), and \( \mathcal{L}_S \) be a probability measure on \( (\mathbb{R}, \mathcal{B}) \).

If

(i) \( S_n \xrightarrow{D,n\to\infty} \mathcal{L}_S \) under \( P_1 \),

(ii) \( P_1 \) is prequentially consistent with respect to \( P_2 \),
then, under suitable smoothness conditions on functions $S_i$ the following result holds:

$$S_n \xrightarrow{D,n \to \infty} L_S \quad \text{under} \quad P_2.$$  

When using the previous conjecture to extract formal sequential Gof tests for plug-in ML SFS's, one will take $P_1 = P^*$ and $P_2 = P_\theta$. Moreover, before using the previous conjecture one needs first to prove Skouras and Dawid conjecture stating that ML plug-in SFS's are sequentially efficient in very great generality.

**Remark** – The previous conjecture could no more be used to deduce omnibus Gof tests if the sequential inference framework is replaced by the standard one where the parameter is estimated once all the data are collected. In the latter case, the probability of $X^{(n)}$ is estimated by $P_{\theta_n, X^{(n)}}$, the one of $X^{(n+1)}$ is estimated by $P_{\theta_{n+1}, X^{(n+1)}}$. But contrary to the sequential case, there is no probability measure on $\Omega$ which is consistent with $P_{\theta_n, X^{(n)}}, P_{\theta_{n+1}, X^{(n+1)}}, \ldots$.

### 4.2 Linking RPIT to the compensator spacings approach

It is easy to see that Arjas and Gasbarra (1997) sequential model fitting approach can be derived from Rosenblatt’s RPIT.

Using the same notations as in sections 2.2 and 3.1, standard results on point processes show that, on one hand:

$$\hat{\Lambda}_T - \hat{\Lambda}_{T_{k-1}} = \int_{T_{k-1}}^{T_k} \hat{\lambda}_s \, ds$$

and on the other hand:

$$F_{k|1,\ldots,k-1}(x) = 1 - \exp \left[ - \int_{T_{k-1}}^{T_{k-1}+x} \hat{\lambda}_s \, ds \right]$$

Then:

$$\forall k \in \mathbb{N}, \quad \tilde{U}_k = F_{k|1,\ldots,k-1}(X_k) = 1 - \exp \left[ - \int_{T_{k-1}}^{T_k} \hat{\lambda}_s \, ds \right] = 1 - \exp \left[ - (\hat{\Lambda}_T - \hat{\Lambda}_{T_{k-1}}) \right]$$

Assessing the predictive performance of competing models with Arjas and Gasbarra compensator spacings approach by testing that $\hat{\Lambda}_{T_{k+1}} - \hat{\Lambda}_{T_k}$ are i.i.d. with exponential distribution under $P^*$ is equivalent to the use of RPIT approach and check the result $\hat{K}_n \xrightarrow{D,n \to \infty} L_{K_0}$ or equivalently $\tilde{U}_k$ are i.i.d. with Uniform$[0,1]$ distribution under $P^*$. So, independence and exponentiality of the compensator spacings are equivalent to the independence and uniformity of the RPIT transformed variables $\tilde{U}_k$ in Rosenblatt result. Prequential model fitting proposed by Arjas and Gasbarra can be considered as a RPIT approach adapted to marked point processes. It checks if “$P \simeq P^*$” but does not give rigorous statistical Gof tests for $H_0 : \{P \in \mathcal{P}\}$. 

9
5 RPIT prequential Gof tests: Numerical investigations

Intensive Monte-Carlo simulations are used to investigate the validity of result (9) under several parametric models: renewal processes, autoregressive processes, non homogeneous Poisson processes, generalized linear models (GLM’s). For each model, m data sets are simulated and $\tilde{K}_{n,p}$ are computed, Edf of the m obtained values of $\tilde{K}_{n,p}$ is compared to the Cdf of $L_{ks}$.

As it will be shown below, experimental results seem to show that, for most models, the sample distribution of $\tilde{K}_{n,p}$ does not depend on the chosen value $\theta_0$ of $\theta$. This is not true for the generalized Pareto distribution and autoregressive models. Moreover, in the major part of the studied models we find that the sample distributions of the computed $\tilde{K}_{n,p}$ are very close to $L_{ks}$. This strongly suggests that for a wide range of models $\tilde{K}_{n,p} \overset{D_{n\infty}}{\rightarrow} L_{ks}$ under $P_\theta$ and consequently that $\tilde{y}_{n,p} \overset{D_{n\infty}}{\rightarrow} \mathbb{B}$ under $P_\theta$.

5.1 General methodology

Monte-Carlo simulations are used here in a four step approach:

1. Choice of a fixed value $\theta_0$ of parameter $\theta$.
2. Simulation of m data sets (of size n) according to the probability distribution $P_{\theta_0}$:
   
   \[ \text{$j^{th}$ data-set: } x_{1}^{(j)}, \ldots, x_{n}^{(j)}, 1 \leq j \leq m \]  

   \( (13) \)

3. For each simulated data-set, the realization $\tilde{k}_{n,p}^{(j)}$ of the r.v. $\tilde{K}_{n,p}$ is computed. Maximum likelihood estimators are used for all the studied models except the renewal GPD model.

4. The Edf of $(\tilde{k}_{n,p}^{(j)})_{1 \leq j \leq m}$ is compared to the Cdf of $L_{ks}$ via the comparison of their upper quantiles denoted respectively $\tilde{q}_\delta$ and $q_k^{ks}$ for orders $\delta \in \{0.50, 0.25, 0.15, 0.10, 0.05, 0.025, 0.01\}$. The mean absolute relative error (MARE):

   \[ \text{MARE} = \sum_{\delta \in \{0.50, \ldots, 0.01\}} \frac{|\tilde{q}_\delta - q_k^{ks}|}{q_k^{ks}} \]

   measures the distance between the Edf of $(\tilde{k}_{n,p}^{(j)})_{1 \leq j \leq m}$ and $L_{ks}$, and therefore the relevance of (9).

Results presented in tables 1, 2, 3 and 4 were obtained with $n = 200$ and $p = 100$. For renewal, Poisson and AR processes, $m = 10.000$ data sets were simulated for each model, when only $m = 1000$ data sets were simulated for ARMA and GLM models. This is due to the fact that for ARMA and GLM models, Splus functions were used for the sequential estimation which were much slower than C language procedures used for the sequential estimation of renewal, Poisson and AR processes.

10
5.2 Renewal processes

Five renewal process models are considered here. The related probability distributions are:

— Gaussian, Exponential (for which result (9) was formally proven) and Uniform.

— Gamma: where the following notations are used: if \( X \) is Gamma\((a, \lambda)\) distributed then, 
\[ E(X) = a \cdot \lambda \text{ and } \text{Var}(X) = a \cdot \lambda^2. \]

— Generalized Pareto distribution (GPD) with scale parameter \( \sigma > 0 \) and shape parameter \( k \).

The Cdf of GPD\((\sigma, k)\), is given by 
\[ G(y;\sigma,k) = 1 - (1 - (\frac{y}{\sigma})^k) \]
if \( k \neq 0 \) and 
\[ G(y;\sigma,k) = 1 - \exp\left(-\frac{y}{\sigma}\right) \]
if \( k = 0 \) where \( y \in \mathbb{R}^+ \) when \( k \leq 0 \), and \( y \in [0, \frac{\sigma}{k}] \) when \( k > 0 \). For \( k \leq -\frac{1}{2} \), \( \text{Var}(\text{GPD}) = \infty \). Probability Weighted Moments (PWM) estimators are used here since they are generally more efficient than ML estimators. The comparative study of Hosking and Wallis (1987) show that PWM estimators have good performance for \(-0.5 < k < 0.5\). These estimators are not defined for \( k = -1 \).

\[
\begin{array}{cccccccccc}
\text{Significance} & \delta & 0.50 & 0.25 & 0.15 & 0.10 & 0.05 & 0.025 & 0.01 & \text{MARE} \\
\hline
\text{Quantiles of } \mathcal{L}_{ks} & 0.822 & 1.019 & 1.138 & 1.224 & 1.358 & 1.480 & 1.628 \\
\text{Gauss}(0,1) & 0.820 & 1.015 & 1.138 & 1.227 & 1.365 & 1.480 & 1.641 \\
\quad (0.2\%) & (0.4\%) & (0.6\%) & (0.2\%) & (0.5\%) & (0.0\%) & (0.8\%) & 0.3\% \\
\text{Expon.}(1) & 0.823 & 1.022 & 1.142 & 1.225 & 1.363 & 1.485 & 1.643 \\
\quad (0.1\%) & (0.3\%) & (0.3\%) & (0.1\%) & (0.4\%) & (0.3\%) & (0.9\%) & 0.3\% \\
\text{Uniform}[2,5] & 0.823 & 1.018 & 1.138 & 1.231 & 1.356 & 1.473 & 1.637 \\
\quad (0.1\%) & (0.1\%) & (0.6\%) & (0.1\%) & (0.5\%) & (0.5\%) & 0.3\% \\
\text{Gamma}(5,2) & 0.814 & 1.010 & 1.130 & 1.217 & 1.349 & 1.473 & 1.627 \\
\quad (1.0\%) & (0.9\%) & (0.7\%) & (0.6\%) & (0.7\%) & (0.5\%) & (0.1\%) & 0.6\% \\
\text{GPD}(0.5,2) & 0.823 & 1.023 & 1.141 & 1.231 & 1.370 & 1.490 & 1.627 \\
\quad (1.3\%) & (0.9\%) & (0.3\%) & (0.6\%) & (0.9\%) & (0.7\%) & (0.1\%) & 0.6\% \\
\text{GPD}(-0.5,4) & 0.834 & 1.032 & 1.151 & 1.238 & 1.374 & 1.490 & 1.650 \\
\quad (1.4\%) & (1.3\%) & (1.1\%) & (1.1\%) & (1.2\%) & (0.7\%) & (1.3\%) & 1.2\% \\
\text{GPD}(-0.9,2) & 0.899 & 1.130 & 1.262 & 1.366 & 1.519 & 1.677 & 1.920 \\
\quad (9.4\%) & (10.9\%) & (11.6\%) & (11.8\%) & (13.3\%) & (17.9\%) & 12.3\% \\
\end{array}
\]

Table 1: Upper quantiles of \( \tilde{K}_{n,p} \) distribution under different renewal processes. Absolute relative errors between upper quantiles from \( \mathcal{L}_{ks} \) and from sample distributions of \( \tilde{K}_{n,p} \) are given between brackets. The mean absolute relative error (MARE) for each line is given in the last column.

Results in table 1 show that the empirical quantiles of \( \tilde{K}_{n,p} \) are very close to those of Kolmogorov-Smirnov \( \mathcal{L}_{ks} \) distribution for almost all the considered renewal processes. This is not the case only for GPD(-0.9,2) where it is known that PWM estimators are not efficient. Further simulation results (see figure 1) show that the mean absolute relative error between
quantiles from $\mathcal{L}_{ks}$ and from sample distributions of $\hat{K}_{n,p}$ generally decrease with $n$. This tends to confirm that (9) holds for a wide range of renewal processes for which the RPIT prequential Gof test could be used.

![Mean Abs. Rel. Error = f(n)](image)

Figure 1: Decrease of mean absolute relative errors (MARE) between quantiles from $\mathcal{L}_{ks}$ and from sample distributions of $\hat{K}_{n,p}$ when $n$ increases.

<table>
<thead>
<tr>
<th>True distribution</th>
<th>Exponential</th>
<th>Uniform</th>
<th>Lognormal</th>
<th>$\chi^2_4$</th>
<th>Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 60$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPIT Test</td>
<td>5.59</td>
<td>60.02</td>
<td>21.92</td>
<td>45.05</td>
<td>43.44</td>
</tr>
<tr>
<td>Standard Test</td>
<td>4.66</td>
<td>96.65</td>
<td>39.99</td>
<td>89.06</td>
<td>88.09</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPIT Test</td>
<td>4.99</td>
<td>86.99</td>
<td>30.55</td>
<td>74.18</td>
<td>70.79</td>
</tr>
<tr>
<td>Standard Test</td>
<td>4.68</td>
<td>99.99</td>
<td>57.63</td>
<td>99.05</td>
<td>98.37</td>
</tr>
<tr>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPIT Test</td>
<td>5.19</td>
<td>99.99</td>
<td>52.69</td>
<td>98.72</td>
<td>97.31</td>
</tr>
<tr>
<td>Standard Test</td>
<td>4.89</td>
<td>100.00</td>
<td>86.61</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 2: Power comparison for the exponential hypothesis: rejection percentage with 5% level tests.

Monte-Carlo simulations show that the RPIT Gof test is slightly less powerful than the specific Gof tests for the different renewal processes. Table 2 gives, for RPIT and standard Kolmogorov-Smirnov Edf exponentiality test (see D’Agostino and Stephens, 1986, pp. 134), the rejection percentages of Exponential distribution when data sets were simulated from the
following distributions: Uniform (on [0,1]), Lognormal (from the standard normal distribution), 
$\chi^2$ and Weibull (with shape parameter $\beta = 1.5$). This table shows that the RPIT test becomes 
nearly as powerful as the standard Edf exponentiality test when $n$ exceeds 200.

5.3 Non homogeneous Poisson processes

For any NHPP model with intensity function $\lambda(t)$, the RPIT test statistics is easily derived using:

$$F_{d|1,...,t-1}(x_t) = 1 - \exp\left[-\int_{t_{t-1}}^{t_t} \lambda(u) \, du\right]. \quad (14)$$

We study here two particular NHPP models:

— Power-Law process (PLP): the non homogeneous Poisson process (NHPP) with intensity 
(for $t > 0$): $\lambda(t) = \alpha \beta^\beta t^{-1}$ where $\alpha$ and $\beta$ are two non-negative parameters. It is proven 
in Crétois et al. (1999) that (9) is true for the PLP.

— The Log-Linear Process (LLP): a NHPP with intensity function $\lambda(t) = \lambda_0 \exp(-\varphi t)$, $\lambda_0 > 0$, $\varphi \in \mathbb{R}$. Results in table 3 suggest that (9) holds also for LLP processes.

<table>
<thead>
<tr>
<th>Significance level $\delta$</th>
<th>0.50</th>
<th>0.25</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>MARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantiles of $L_{ks}$</td>
<td>0.822</td>
<td>1.019</td>
<td>1.138</td>
<td>1.224</td>
<td>1.358</td>
<td>1.480</td>
<td>1.628</td>
<td></td>
</tr>
<tr>
<td>PLP</td>
<td>0.825</td>
<td>1.019</td>
<td>1.138</td>
<td>1.227</td>
<td>1.363</td>
<td>1.480</td>
<td>1.643</td>
<td></td>
</tr>
<tr>
<td>(0.4%) (0.0%) (0.0%) (0.2%)</td>
<td>(0.4%) (0.0%) (0.0%) (0.0%) (0.9%) (0.9%)</td>
<td>0.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLP</td>
<td>0.829</td>
<td>1.028</td>
<td>1.150</td>
<td>1.233</td>
<td>1.374</td>
<td>1.494</td>
<td>1.649</td>
<td></td>
</tr>
<tr>
<td>(0.8%) (0.9%) (1.0%) (1.0%)</td>
<td>(0.9%) (0.7%) (1.2%) (0.9%) (1.3%) (1.3%)</td>
<td>1.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Upper quantiles of $\tilde{K}_{n,p}$ distribution under different NHPP models. Absolute relative errors between upper quantiles from $L_{ks}$ and from sample distributions of $\tilde{K}_{n,p}$ are given between brackets. The mean absolute relative error (MARE) for each line is given in the last column.

In Crétois et al. (1999) the power of the RPIT Gof test for testing the PLP was compared to 
two classical PLP Gof tests (namely the Rigdon and TTT-plot tests). It was shown that, even 
if it is not as powerful as the specific PLP tests, RPIT Gof test performs very well, especially 
for small samples.

5.4 Autoregressive processes

Here, prequential Gof tests are numerically studied for the following processes:

— AR(1) where for any integer $i > 0$: $X_i = \phi.X_{i-1} + \varepsilon_i$, 

13
- AR(2) where \( X_i = \phi_1 X_{i-1} + \phi_2 X_{i-2} + \epsilon_i \),

- ARMA(1,1) where \( X_i - \phi_1 X_{i-1} = \epsilon_i + \theta_1 \epsilon_{i-1} \)

Disturbances \( \epsilon_i \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \).

As it was mentioned by Skouras and Dawid (1999), when \( |\phi| > 1 \) the AR(1) process is explosive and there is no efficient plug-in SFS. On the other hand, when \( |\phi| < 1 \), the process is stationary and any plug-in SFS based on ML, least squares or any ridge estimator is efficient. AR(2) processes are stationary when \( |\phi_2| < 1 \), \( \phi_1 + \phi_2 < 1 \) and \( \phi_2 - \phi_1 < 1 \).

<table>
<thead>
<tr>
<th>Significance level ( \delta )</th>
<th>0.50</th>
<th>0.25</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>MARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantiles of ( L_{ks} )</td>
<td>0.822</td>
<td>1.019</td>
<td>1.138</td>
<td>1.224</td>
<td>1.358</td>
<td>1.480</td>
<td>1.628</td>
<td></td>
</tr>
<tr>
<td>AR(1): ( \phi = 0.5 ) ( \sigma = 1.0 ) Station.</td>
<td>(0.6%)</td>
<td>(0.1%)</td>
<td>(0.2%)</td>
<td>(0.1%)</td>
<td>(0.7%)</td>
<td>(0.5%)</td>
<td>(1.3%)</td>
<td>0.5%</td>
</tr>
<tr>
<td>AR(1): ( \phi = -0.9 ) ( \sigma = 3.0 ) Station.</td>
<td>(0.1%)</td>
<td>(0.1%)</td>
<td>(0.3%)</td>
<td>(0.0%)</td>
<td>(0.2%)</td>
<td>(1.1%)</td>
<td>(1.7%)</td>
<td>0.5%</td>
</tr>
<tr>
<td>AR(1): ( \phi = 1.1 ) ( \sigma = 1.0 ) Non-Station.</td>
<td>(11.5%)</td>
<td>(10.4%)</td>
<td>(10.3%)</td>
<td>(9.8%)</td>
<td>(9.2%)</td>
<td>(8.4%)</td>
<td>(8.8%)</td>
<td>9.8%</td>
</tr>
<tr>
<td>AR(2): ( \phi_1 = 0.2 ) ( \phi_2 = 0.5, \sigma = 1.0 ) Stationary</td>
<td>(0.1%)</td>
<td>(0.3%)</td>
<td>(0.6%)</td>
<td>(0.2%)</td>
<td>(0.8%)</td>
<td>(0.5%)</td>
<td>(0.4%)</td>
<td>0.4%</td>
</tr>
<tr>
<td>AR(2): ( \phi_1 = -0.5 ) ( \phi_2 = 0.3, \sigma = 2.0 ) Stationary</td>
<td>(0.7%)</td>
<td>(0.3%)</td>
<td>(0.4%)</td>
<td>(0.7%)</td>
<td>(0.6%)</td>
<td>(0.9%)</td>
<td>(1.5%)</td>
<td>0.7%</td>
</tr>
<tr>
<td>ARMA(1,1): ( \phi_1 = 0.5 ) ( \theta_1 = 0.4, \sigma = 1.0 ) Stationary</td>
<td>(3.2%)</td>
<td>(3.0%)</td>
<td>(1.7%)</td>
<td>(1.1%)</td>
<td>(1.6%)</td>
<td>(2.3%)</td>
<td>(3.7%)</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Table 4: Upper quantiles of \( \tilde{K}_{np} \) distribution under different auto-regressive processes. Absolute relative errors between upper quantiles from \( L_{ks} \) and from sample distributions of \( \tilde{K}_{np} \) are given between brackets. The mean absolute relative error (MARE) for each line is given in the last column.

Using RPIT Gof test for AR processes needs only to know the conditional distribution of \( X_i \) given \( [X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \) under \( P_\theta \). This is \( \mathcal{N}(\phi X_{i-1}, \sigma^2) \) for AR(1) and \( \mathcal{N}(\phi_1 X_{i-1} + \phi_2 x_{i-2}, \sigma^2) \) for AR(2). RPIT prequential Gof test is more difficult to use for validating general
ARMA\((p,q)\) models where

\[
X_i - \phi_1 X_{i-1} - \ldots - \phi_p X_{i-p} = \epsilon_i + \theta_1 \epsilon_{i-1} + \ldots + \theta_q \epsilon_{i-q}
\text{ with } \epsilon_i \text{ i.i.d. } \mathcal{N}(0, \sigma^2)
\]

since conditional distributions of \(X_i\) given \(X_1, \ldots, X_{i-1}\) depend on the non-observed past disturbances \(\epsilon_{i-1}, \epsilon_{i-2}, \ldots\). Nevertheless, when the ARMA process \(X\) is invertible, \(\epsilon_{i-1}, \epsilon_{i-2}, \ldots\) can be deduced from past observations \(x_{i-1}, x_{i-2}, \ldots\) (see for example Brockwell and Davis, 1996). Therefore, one can deduce the conditional distribution of \(X_i\) given \([X_1 = x_1, \ldots, X_{i-1} = x_{i-1}]\) which is \(\mathcal{N}(\phi_1 x_{i-1} + \ldots + \phi_p x_{i-p} + \theta_1 \epsilon_{i-1} + \ldots \theta_q \epsilon_{i-q}, \sigma^2)\).

According to results of table 4, conjecture (9) seems to hold (\(\tilde{K}_{n,p}\) is approximately \(\mathcal{L}_{k,n}\) distributed) for stationary AR(1) and AR(2) processes. This is not true for the explosive process simulated. Results for ARMA(1,1) process need further investigations, even if a MARE of 2.4\% tends to confirm conjecture (9).

To assess the power of RPIT Gof test, blocks of 10,000 data sets were simulated for different AR(2) models. For each data set, AR(1) was tested using RPIT Gof test. Results given in table 5 show that the power of RPIT AR(1) test when AR(2) is simulated increases with non-negative values of \(\phi_2\). But RPIT Gof test seems to be unable to reject AR(1) assumption when AR(2) is simulated with \(\phi_2 < 0\). RPIT Gof test also detects non stationarity since AR(1) was rejected for all data sets simulated with a non stationary AR(1) (see column \(\phi_2 = 0, \phi_1 = 1.2\) marked with * in table 5.

\[
\begin{array}{cccccccc}
\phi_2 & -0.9 & -0.5 & -0.3 & 0 & 0^* & 0.3 & 0.5 & 0.9 \\
\phi_1 & 1 & 1 & 1 & 0.5 & 1.2^* & -0.5 & 0.2 & 0 \\
\hline
\text{% of AR(1)} & & & & & & & & \\
\text{rejection} & 8.3 & 1.0 & 1.9 & 5.1 & 100.0^* & 12.8 & 22.5 & 71.0 \\
\end{array}
\]

Table 5: Power of RPIT Gof test when testing AR(1) on AR(2) simulated data.

### 5.5 Generalized Linear Models

Selecting the appropriate GLM for a given data set is an actual research problem. Several recent papers give statistical tools to select the right distribution, link function and covariates (see for example Paul and Deng, 2000, and Qian et al., 1996). RPIT Gof test, as shown hereafter, could be an efficient tool to select the right GLM.

Three GLM's are considered here:

- Gaussian linear model where \(X_i\)'s are independent and \(X_i \sim \mathcal{N}(\beta_0 + \beta_1 Y_i^{(1)} + \beta_2 Y_i^{(2)}, \sigma^2)\), with two given covariates \(Y^{(1)}\) and \(Y^{(2)}\).
— Gamma inverse-link GLM where \( X_i \)’s are independent and \( X_i \sim \text{Gamma}(a = 2, \lambda_i) \) with 
\[
\mathbb{E}(X_i) = a \cdot \lambda_i = \left( \beta_0 + \beta_1 Y_i^{(1)} + \beta_2 Y_i^{(2)} \right)^{-1}.
\]

— Exponential log-link GLM where \( X_i \)’s are independent and \( X_i \sim \text{Exponential}(\lambda_i) \) with 
\[
\mathbb{E}(X_i) = \lambda_i^{-1} = \beta_0 \exp(\beta_1 i). \] This is the Moranda geometric model (see Moranda, 1975) 
used for the analysis of repairable systems reliability.

<table>
<thead>
<tr>
<th>Significance level ( \delta )</th>
<th>0.50</th>
<th>0.25</th>
<th>0.15</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>MARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantiles of ( L_{ks} )</td>
<td>0.822</td>
<td>1.019</td>
<td>1.138</td>
<td>1.224</td>
<td>1.358</td>
<td>1.480</td>
<td>1.628</td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.819</td>
<td>1.005</td>
<td>1.128</td>
<td>1.233</td>
<td>1.368</td>
<td>1.478</td>
<td>1.590</td>
<td></td>
</tr>
<tr>
<td>identity-link</td>
<td>(0.4%)</td>
<td>(1.4%)</td>
<td>(0.9%)</td>
<td>(0.7%)</td>
<td>(0.7%)</td>
<td>(0.1%)</td>
<td>(2.3%)</td>
<td>0.9%</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.823</td>
<td>1.033</td>
<td>1.134</td>
<td>1.214</td>
<td>1.349</td>
<td>1.480</td>
<td>1.563</td>
<td></td>
</tr>
<tr>
<td>inverse-link</td>
<td>(0.1%)</td>
<td>(1.4%)</td>
<td>(0.3%)</td>
<td>(0.8%)</td>
<td>(0.7%)</td>
<td>(0.0%)</td>
<td>(4.0%)</td>
<td>1.0%</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.822</td>
<td>1.019</td>
<td>1.138</td>
<td>1.226</td>
<td>1.360</td>
<td>1.491</td>
<td>1.642</td>
<td></td>
</tr>
<tr>
<td>log-link</td>
<td>(0.0%)</td>
<td>(0.0%)</td>
<td>(0.0%)</td>
<td>(0.2%)</td>
<td>(0.1%)</td>
<td>(0.7%)</td>
<td>(0.9%)</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Table 6: Upper quantiles of \( \tilde{K}_{n,p} \) distribution under different GLM’s. Absolute relative errors 
between upper quantiles from \( L_{ks} \) and from sample distributions of \( \tilde{K}_{n,p} \) are given between 
brackets. The mean absolute relative error (MARE) for each line is given in the last column.

Table 6 suggests that the RPIT prequential Gof test can also be used to choose between 
competing GLM’s since sample distributions of \( \tilde{K}_{n,p} \) are still very close to \( L_{ks} \).

It was not possible to study the power of RPIT Gof test when wrong GLM’s were tested 
since when the distribution or the link function of the tested GLM was misspecified, \texttt{Splus} GLM 
algorithm based on iterative weighted least squares did not converge.

6 Discussion and open questions

We investigated in this work the use of RPIT criterion as an omnibus Gof test for a wide 
range of stochastic process models. This Gof test, based on the plug-in prequential approach, is 
conceptually close to Seillier-Moiseiwitsch et al. (1992) approach based on a martingale central 
limit theorem, except that, in their approach, one should first check Lindeberg type conditions. 
Even if the conjecture decribed by (9), which represents the theoretical base of our omnibus 
Gof test, is formally proved only for some Poisson processes, extended numerical experiments 
suggest that this conjecture holds for many other stochastic process models: renewal processes, 
GLM’s, stationnary autoregressive processes, etc. This justifies the use of RPIT criterion as an
omnibus GoF test in a finite sample context. Simulations show also that the loss of power of RPIT GoF test in comparison to model-specific GoF tests is not dramatic. RPIT approach can not be used as a statistical GoF test when there are no efficient estimators; this is the case for explosive AR processes, and GPD distributions where the shape parameter $k$ is less than $-1/2$. Besides the practical consequences of this study, i.e. the possible use of RPIT as a GoF test for a wide range of process families, many questions remain open and need further theoretical investigations:

- Does Dawid’s conjecture concerning the prequential efficiency of Plug-in ML SFS hold?
- What argument (efficiency or consistency) makes the empirical distribution of $\tilde{K}_{n,p}$ so close to $L_{ka}$?
- How to choose optimal values of $p$ (number of observations used in the first estimation step)?
- How this approach could be adopted in non parametric models (cf. discussion in Skouras and Dawid 1999)?
- How RPIT GoF test could be related to the Predictive Minimum Description Length tool especially for validating Time Series models (see Hansen and Yu 2001)?

References


