

Functional approach to cluster municipalities with respect to air quality assessment

Rosaria Ignaccolo¹, Stefania Ghigo¹, and Stefano Bande²

¹ Dipartimento di Statistica e Matematica Applicata “Diego de Castro”
Università degli studi di Torino, Torino, Italy
(e-mail: ignaccolo@econ.unito.it, ghigo@econ.unito.it)

² ARPA Piemonte, Torino, Italy
(e-mail: s.bande@arpa.piemonte.it)

Extended Abstract

In this work, we propose a functional approach to cluster municipalities into zones characterized by different criticality levels of atmospheric pollution. Specifically, we consider air pollutant time series as functional data ([9]) and propose strategies to cluster them as well as to aggregate them at municipality scale. Our proposal is meant to meet a request of the European Air Quality Framework Directive (1996/62/CE) that imposes land classification, so-called “zoning”, to distinguish zones which need further actions from those which only need maintenance. The national implementations of EU Directives delegate the air quality assessment responsibility to Italian Regions and here our proposed approach is applied to Piemonte (Northern Italy) air quality data of the year 2005.

Data are output of a three-dimensional deterministic modeling system implemented by the *Area Previsione e Monitoraggio Ambientale* of the environmental agency *ARPA Piemonte*. Main atmospheric pollutant concentration fields (such as CO, SO₂, PM₁₀ and NO₂) are produced on an hourly basis over a regular grid that has a horizontal resolution of 4 km and covers Piemonte, neighboring Italian regions and foreign countries [2]. Hence available data are not observed but “artificial”, however the European law allows their use in the air quality assessment. Nevertheless, comparing the output of the deterministic model with observations provided by the Piemonte monitoring network it turns out that concentrations are sometime underevaluated or overevaluated. Hence, a procedure of kriging with external drift is used to assimilate observed data in concentration fields (see [10] and [11]), that are “corrected” before to be used in the clustering. A cross validation analysis shows that kriging results are satisfactory (for further details see Ghigo (2009) [6]).

Since municipalities are the reference territorial administrative units for undertaking actions, we propose to upscale data from a regular grid to municipalities. The aggregation at municipality scale can be realized solving the so-called “change of support problem”, retrieving a value of a random field

on an area starting from its values on points ([5] and [4]) for every fixed time. The first simple solution is to integrate over the area, that means to average the field values weighted by areas over the cells belonging to a certain municipality (*MeanAreas*). Other two alternatives for the aggregation are the average of the field values weighted by the building percentage (*MeanBuilt*) for every cell - a point in the grid represents a cell - and the 90th percentile over the cell values in a municipality (*90perc*). The building percentage is an important indicator of the anthropic activity which could generate more pollution in a municipality, whereas a wide country area could not contribute at all. Instead the 90th percentile is chosen as a measure of extreme cases in a municipality, in a precautionary perspective.

The pre-processing of data and the municipality upscaling procedures are applied singularly for each pollutant. In order to look at the global air quality status and to obtain an overall Piemonte zoning, we propose two strategies to summarize time series (getting pollutants aggregation): evaluating an air quality index introduced by Bruno and Cocchi [3] and carrying out a Multivariate Functional Principal Component Analysis (MFPCA).

Within the BC (Bruno and Cocchi) air quality family an aggregation over pollutants by the maximum function is proposed, in order to keep information about critical cases, to obtain BC index time series for all the municipalities. Considering these time series as functional data, and preserving their temporal patterns, we can cluster them ([1]) and obtain groups of municipalities, through a functional cluster analysis where Partitioning Around Medoids algorithm (PAM, [8]) is embedded (see also [7]). PAM allows to have an object - the so-called “medoid” - representing the cluster, which in this case will be a curve showing the temporal evolution of the air quality index.

As alternative technique, we explore the Functional Principal Component Analysis in its multivariate version that allow to consider several pollutants taking into account their interactions. At a second step, we apply the PAM algorithm to the scores of the principal components, obtaining groups of municipalities. In this case the medoids are scores where the temporal component is integrated out.

Results obtained by the proposed methodologies are displayed in Figure 1. Using color gradations (as traffic lights) changing with cluster concentration mean values allows to better identify different criticality levels. The most critical group, the red one, lengthens over the main road network: it includes big towns - as Torino, Alessandria, Novara - and their suburbs, characterized by industrialization. The orange cluster is formed by piedmont municipalities, whereas mountain municipalities are grouped in the green cluster, that is the less critical one.

Looking at the six maps it seems that the municipalities clusters obtained by the two different summary methods and the three different upscaling algorithms are quite similar. The maps (a)-(c) in Figure 1 are the result of multi-pollutant zoning through BC index: the small number of municipalities

migrating from a group to another one when we change the upscaling algorithm confirms that these maps are not so different. However, the functional clustering on BC index time series allows us to look at the representative temporal evolution in the different zones by medoids.

In the two block average upscaling cases with MFPCA clustering (maps (d) and (e) in Figure 1), the obtained groups of municipalities are slightly different than in the 90th percentile upscaling case (map (f)). This difference seems due to the importance of SO_2 in the second principal component when we carry out the MFPCA with *MeanAreas* and *MeanBuilt*, while with *90perc* the variability is mostly explained by PM_{10} .

All the results can be considered precautionary. We will furtherly discuss and compare the results of the different analysis strategies. For this goal, we also quantify the migration of municipalities among clusters and visualize it constructing three-color maps of differences. On another hand, policy makers could choose to adopt one of our proposals as the best strategy taking into account their knowledge about the land use and their constraints in making decisions for the different zones.

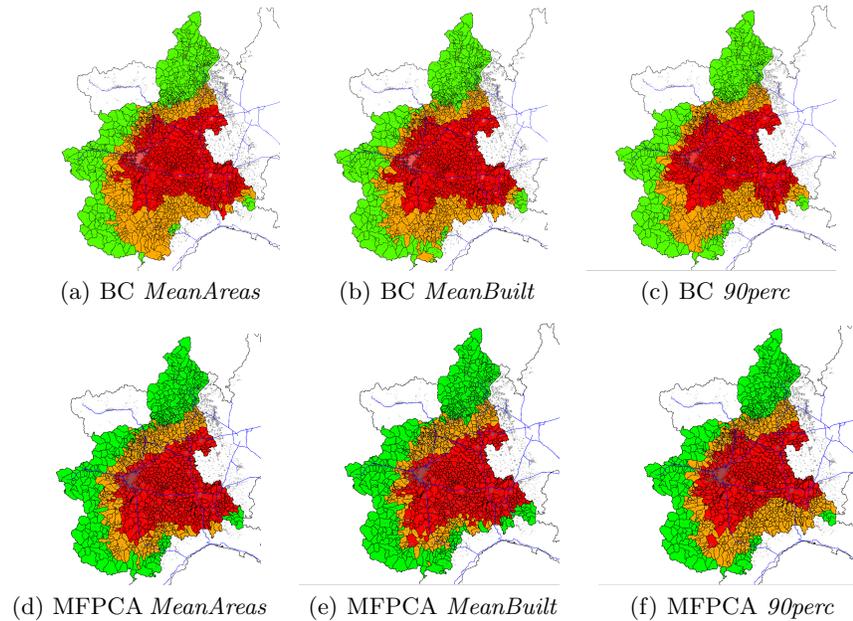


Fig. 1. Multi-pollutant risk maps obtained through BC index (top) and MFPCA (bottom). Traffic light colors are associated with criticality of air quality status.

References

1. Abraham, C., Cornillon, P. A., Matzner-Løber, E., Molinari, N. “Unsupervised curve clustering using B-splines”. *Scand. J. Statist.* 30, 581–595 (2003)
2. Bande, S., Clemente, M., De Maria, R., Muraro, M., Piccolo, M., Arduino, G., Calori, G., Finardi, S., Radice, P., Silibello, C., Brusasca, G. “The modelling system supporting Piemonte region yearly air quality assessment”. *Proceedings of the 6th International Conference on Urban Air Quality* (2007)
3. Bruno, F., Cocchi, D. “A unified strategy for building simple air quality indexes”. *Environmetrics* 13, 243–261 (2002)
4. Chiles, J.P., Delfiner, P. *Geostatistics. Modeling Spatial Uncertainty*. John Wiley and Sons (1999)
5. Cressie, N. *Statistics for Spatial Data*. John Wiley and Sons (1993)
6. Ghigo, S. “Metodi di statistica funzionale a supporto del monitoraggio e della valutazione della qualità dell’aria”. *Tesi di Dottorato*, Università degli Studi di Torino (2009)
7. Ignaccolo, R., Ghigo, S., Giovenali, E. “Analysis of air quality monitoring networks by Functional Clustering”. *Environmetrics* 19, 243–261 (2008)
8. Kaufman, L., Rousseeuw, P. J. *Finding Groups in Data. An introduction to cluster analysis*. Wiley (1990)
9. Ramsay, J. O., Silverman, B. W. *Functional data analysis*. Springer (2005)
10. Van de Kasstele, J., Stein, A. and Dekkers, A.L.M., Velders, G.J.M. “External drift kriging of NO_x concentrations with dispersion model output in a reduced air quality monitoring network”. *Environmental and Ecological Statistics* 16(3), 321–339 (2009)
11. Wackernagel, H. *Multivariate geostatistics: an introduction with applications*. Springer (2003)

Multiple Model Adaptive Control with Mixing: Discrete-time Case

Petros Ioannou¹, Edoardo Mosca², and Simone Baldi²

¹ Department of Electrical Engineering
University of Southern California, Los Angeles, CA, USA
(e-mail: ioannou@usc.edu)

² Dipartimento Sistemi e Informatica
Università di Firenze, 50139 Firenze, Italy
(e-mail: mosca@dsi.unifi.it, sbaldi@dsi.unifi.it)

Abstract. The main result of this paper is the extension of the AMC approach exposed in [1], to the discrete-time setting. Besides, extension to the tracking problem is considered. The stability and robustness properties of the adaptive mixing control scheme are analyzed. It is shown that in the ideal case, when no disturbances or unmodelled dynamics are present, the tracking error converges to zero; otherwise the mean-square tracking error is of the order of the modeling error provided the unmodeled dynamics satisfy a norm-bound condition.

Keywords: Robust adaptive control, Multiple model adaptive control.

1 Introduction

All real systems are subjected to uncertainty due to unmodeled dynamics, unknown system parameters, disturbances, and process changes. A practical control design must be able to maintain performance and stability robustly in the presence of these uncertainties. When model uncertainties are sufficiently small, modern linear time invariant (LTI) control theories, *e.g.*, \mathcal{H}_∞ and μ -synthesis, ensure satisfactory closed-loop objectives. However, changes in operating conditions, failure or degradation of components, or unexpected changes typically lead to a large parametric uncertainty, with the result that a single fixed LTI controller may no longer achieve satisfactory closed-loop behavior, let alone stability. The approach taken in this work is a novel multiple model adaptive control approach called *adaptive mixing control* (AMC), shown in Fig. 1. Each of the N candidate controllers C_1, \dots, C_N is tuned to a small subset of the parameter uncertainty. The set of candidate controllers is sufficiently rich such that for every admissible plant there exists at least one controller that achieves the performance objective. By monitoring the plant's input/output data, the robust adaptive supervisor system 'mixes' the candidate controllers. The supervisor comprises two subsystems: the online parameter estimator and the mixer. The online parameter estimator generates real-time estimates $\theta(t)$ of the unknown parameter vector θ^* , and the mixer determines the participation level each candidate controller

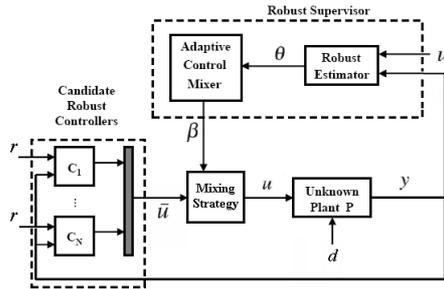


Fig. 1. AMC architecture

based on $\theta(t)$. The AMC approach, developed in a continuous-time setting in [1], is here analyzed in discrete-time: we also extend the adaptive control objective to include the tracking problem, whereas [1] dealt with the regulation problem. We establish that the closed-loop states remain bounded when the scheme is applied to the true plant with multiplicative uncertainty and bounded disturbance. When the true system matches the nominal model and in the absence of an external disturbance, the tracking error e_1 and input u_p converge to zero.

The paper is organized as follows: preliminary definition and notation are exposed in section II. Section III deals with the problem formulation and the main theorem. The key results used in the stability and robustness analysis of the adaptive mixing control scheme are stated in the appendix (the proofs are omitted for lack of space).

2 Notation and Preliminaries

For $A \in \mathbb{R}^{m \times n}$, the transpose of A is denoted by A^T . If $y : \mathbb{Z}^+ \rightarrow \mathbb{R}^n$, then the $l_{2\delta}$ norm of y is:

$$\|y_k\|_{2\delta} \triangleq \left(\sum_{i=0}^k \delta^{k-i} y^T(i)y(i) \right)^{1/2} \quad (1)$$

where $0 < \delta \leq 1$ is a constant, provided that the sum in (1) exists. By $\|y_k\|_2$ we mean $\|y_k\|_{2\delta}$ with $\delta = 1$, and we say that $y \in l_{2e}$ if $\|y_k\|_2$ exists $\forall k \in \mathbb{Z}^+$. Let $y \in l_{2e}$, and consider the set

$$\mathcal{S}(\mu) = \left\{ y : \sum_{i=k}^{k+N-1} y^T(i)y(i) \leq c_0\mu N + c_1, \quad \forall k, N \geq 0 \right\} \quad (2)$$

for a given constant μ , where $c_0, c_1 \geq 0$ are some finite constants independent of μ . We say that y is μ -small in the mean square sense (m.s.s.) if $y \in \mathcal{S}(\mu)$. Let $H(z)$ and $h(k)$ be the transfer function and impulse response, respectively,

of some LTI system. If $H(z)$ is a proper transfer function and analytic in $|z| \geq \sqrt{\delta}$ for some $\delta \geq 0$, then the $\|\cdot\|_{2\delta}$ system norm of $H(z)$ is defined as $\|H\|_{2\delta} \triangleq \frac{1}{\sqrt{2\pi}} \left\{ \int_0^{2\pi} |H(\sqrt{\delta}e^{j\omega})|^2 d\omega \right\}^{1/2}$. The induced l_∞ system norm of H is given by $\|H\|_1 \triangleq \|h\|_1$.

3 Problem Formulation

The objective is to design a controller for the uncertain plant

$$y_p = G(z, \theta^*)(u_p + d), \quad G(z, \theta^*) = G_0(z, \theta^*)(1 + \Delta_m(z)) \quad (3)$$

$$G_0(z, \theta^*) = \frac{N_0(z, \theta^*)}{D_0(z, \theta^*)} = \frac{\theta_b^{*T} \alpha_m(z)}{z^n + \theta_a^{*T} \alpha_{n-1}(z)} \quad (4)$$

where $G_0(z, \theta^*)$ represents the nominal plant; the vector $\theta^* \triangleq [\theta_b^{*T} \ \theta_a^{*T}]^T \in \Omega \subset \mathbb{R}^{2n}$ contains the unknown parameters of $G_0(z, \theta^*)$; $\alpha_{n-1}(z) \triangleq [z^{n-1} \ z^{n-2} \ \dots \ z \ 1]$; $\Delta_m(z)$ is an unknown multiplicative perturbation; and d is a bounded disturbance, i.e., $|d(k)| \leq d_0, \forall k \in \mathbb{Z}_+$. We make the following plant assumptions:

- P1. $D_0(z, \theta^*)$ is a monic polynomial whose degree n is known.
- P2. Degree $(N_0(z, \theta^*)) = m \leq n - 1$.
- P3. $\Delta_m(z)$ is proper, rational, and analytic in $|z| \geq \sqrt{\delta_0}$ for some known $\delta_0 > 0$.
- P4. $\theta^* \in \Omega$ for some known compact convex set $\Omega \subset \mathbb{R}^{2n}$.

It should be emphasized that both unstable and nonminimum phase plants are admissible despite requirements P1-P4. The control objective is to choose the plant input u_p so that the plant output y_p follow a certain class of reference signals y_m . We include tracking by using the internal model principle as follows: the reference signal $y_m \in l_\infty$ is assumed to satisfy

$$Q_m(z)y_m(k) = 0 \quad (5)$$

where $Q_m(z)$, known as the internal model of y_m , is a known monic polynomial of degree q with all roots in $|z| \leq 1$. $Q_m(z)$ is assumed to satisfy

- P5. $Q_m(z), Z_p(z)$ are coprime.

The parameter set Ω is divided into N smaller subsets $\Omega_1, \dots, \Omega_N$. The parameter partition $\mathcal{P} \triangleq \{\Omega_i \subset \mathbb{R}^{2n}\}_{i \in \mathcal{I}}$, where \mathcal{I} denotes the index set $\{1, \dots, N\}$ is developed such that each parameter subset Ω_i is compact and \mathcal{P} covers Ω , i.e., $\Omega \subset \cup_{i \in \mathcal{I}} \Omega_i$. For each subset Ω_i a LTI controller with rational transfer function $C_i(z)$ is synthesized such that the control law $u_p = -C_i(z)(y_p - y_m)$ yields a stable closed-loop system that meets the performance requirements in the subset Ω_i . Given the family of N candidate

controllers $\mathcal{C} \triangleq \{C_i(z)\}_{i \in \mathcal{I}}$, a multicontroller $\mathbb{C}(\beta)$ is constructed from \mathcal{C} . The multicontroller is a dynamical system capable of generating each candidate control laws, as well as a mix of candidate control laws. The multicontroller depends on the mixing signal $\beta = [\beta_1, \dots, \beta_N]^T \in [0, 1]^N$ which determines the participation level of the candidate controllers. For fixed values of β the multicontroller $u_p = -C(z; \beta)(y_p - y_m)$ has the transfer function:

$$\frac{\hat{P}(z; \beta)}{\hat{L}(z; \beta)Q_m(z)} = \frac{p_0(\beta)z^{r+q-1} + \bar{p}^T(\beta)\alpha_{r+q-2}(z)}{z^{r+q-1} + \bar{l}^T(\beta)\alpha_{r+q-2}(z)} \quad (6)$$

The mixer implements the mapping $\beta : \Omega \mapsto \mathcal{B}_\theta \subset [0, 1]^N$. The following property of β and of the multicontroller $\mathbb{C}(\beta)$ are assumed

- M1. $\beta(\theta)$ is Lipschitz in Ω .
- C1. The elements $p_0(\beta)$, $\bar{p}(\beta)$, and $\bar{l}(\beta)$ are Lipschitz respect to β .
- C2. For all $\theta^* \in \Omega$, let $\beta^* \triangleq \beta(\theta^*)$; then $C(z; \beta^*)$ internally stabilizes the plant $G(z; \theta^*)$.

Property M1, together with C1 ensures that if θ is tuned slowly then the closed-loop system will vary slowly. Property C2 ensures that $\mathbb{C}(\beta(\theta))$ is a certainty equivalence stabilizing controller. Construction of the multicontroller involves interpolating the candidate controllers over the parameter overlaps. Numerous controller interpolation approaches have been proposed in the context of gain scheduling. These methods interpolate controller poles, zeros, and gains [2]; solutions of the Riccati equations for an H_∞ design [3]; state and observer gains [4]; controller output blending, *i.e.*, $u = \sum_{i=1}^p u_i$. As in gain scheduling, interpolation methods may not satisfy the point-wise stability requirement C2 (cf. the counter example of [4]) that should be previously verified. Otherwise, there also exist theoretically justified methods, which can be used to construct the multicontroller in order to assure property C2 [4].

The adaptive mixing law approach replaces θ^* with its estimate θ_p . Because of the presence of multiplicative uncertainty $\Delta_m(z)$ and disturbance d , we use a *robust online parameter estimator*.

$$\theta_p(k) = \text{Pr}_\Omega(\theta_p(k-1) + \Gamma \epsilon(k)\phi(k)), \quad \epsilon(k) = \frac{z(k) - \theta_p^T(k-1)\phi(k)}{m_s^2(k)} \quad (7)$$

$$m_s^2(k) = 1 + \phi^T(k)\phi(k) + n_d(k), \quad n_d(k+1) = \delta_0 n_d(k) + u_p^2(k) + y_p^2(k) \quad (8)$$

where Pr stands for the projection operator that forces the estimated parameters to stay within a specified convex set in the parameter space, $0 < \Gamma < 2$, $z(k) = \frac{z^n}{A_p(z)} F_\eta(z) y_p(k)$, $\phi(k) = \left[\frac{\alpha_m^T(z)}{A_p(z)} F_\eta(z) u_p(k) \quad - \frac{\alpha_{n-1}^T(z)}{A_p(z)} F_\eta(z) y_p(k) \right]^T$, A_p is a Hurwitz polynomial of degree n , and F_η a proper stable minimum-phase filter. The adaptive law (7)-(8) guarantees:

- E1. $\epsilon(k)$, $\epsilon(k)m_s(k)$, $\theta_p(k) \in l_\infty$.

- E2. $\epsilon(k), \epsilon(k)m_s(k), |\theta_p(k) - \theta_p(k-1)| \in \mathcal{S}\left(\frac{\eta^2}{m_s^2}\right)$ if $\Delta_m, d \neq 0$.
 E3. $\epsilon(k), \epsilon(k)m_s(k), |\theta_p(k) - \theta_p(k-1)| \in l_2$ if $\Delta_m, d = 0$.

Theorem 1. *Let the unknown plant be given by (3)-(4) with internal model (5) and satisfying the plant assumptions P1-P5. Consider the adaptive mixing controller with the multicontroller $\mathbb{C}(\beta(\theta))$ given by (6) and satisfying assumptions C1-C2; mixer satisfying M1; and robust adaptive law given by (7)-(8). Then the following results hold*

1. If $\Delta_m, d = 0$, then $u_p(k), e_1(k) \triangleq y_p(k) - y_m(k) \rightarrow 0$ as $k \rightarrow \infty$.
2. If $\Delta_m, d \neq 0$, there exists $\mu^* > 0$ such that, if $c(\Delta_1^2 + \Delta_2^2) < \mu^*$ where $\Delta_1 = \left\| \frac{N_0 \Delta_m F_\eta}{\Lambda_p} \right\|_{2\delta_0}, \Delta_2 = \left\| \frac{N_0(1+\Delta_m)F_\eta}{\Lambda_p} \right\|_1 d_0$ and $c > 0$ a finite constant, then the adaptive mixing control scheme guarantees $u_p, e_1, \Delta u_p, \Delta e_1 \in l_\infty$ and $\sum_{i=0}^{k-1} |e_1(i)|^2 \leq c_0 \mu^2 k + c_1$, where $\mu^2 = c(\Delta_1^2 + \Delta_2^2)$.

Proof - For lack of space, only the main points of the proof are given. Case (1):

Step 1. Following the guidelines used in [5] for the stability proof of the discrete-time Adaptive Pole Placement Control (APPC) we can manipulate the control law and the normalized estimation error equations to obtain

$$x(k+1) = A(k)x(k) + b_1(k)\epsilon m_s^2(k) + b_2 y_{m1}(k) \quad (9)$$

$$u_p(k) = C_1^T(k)x(k) + d_{11}(k)\epsilon m_s^2(k) + d_{12}y_{m1}(k) \quad (10)$$

$$y_p(k) = C_2^T(k)x(k) + d_{21}(k)\epsilon m_s^2(k) + d_{22}y_{m1}(k) \quad (11)$$

where $x \triangleq [y_f(k + \bar{n} + q - 2), \dots, y_f(k), u_f(k + \bar{n} + q - 2), \dots, u_f(k)]^T$, $u_f = u/\Lambda$, $y_f = y/\Lambda$, Λ a Hurwitz polynomial, $\bar{n} = \max\{n, r\}$, $y_{m1} \triangleq \frac{\hat{P}}{\Lambda} y_m \in l_\infty$.

Step 2. We will establish the exponential stability of the homogeneous part of (9). It can be demonstrated that for each frozen time k

$$\det(zI - A(k)) = \hat{R}_p \bar{\Lambda}_q \cdot \hat{L} Q_m + \hat{P} \hat{Z}_p \bar{\Lambda}_q = A^*(z, k) \bar{\Lambda}_q(z) \quad (12)$$

where $A^*(z, k)$ is the characteristic polynomial of the closed-loop formed by the estimated plant and the controller. $\bar{\Lambda}_q$ is a Hurwitz polynomial, and thanks to C2, $A^*(z, k)$ is Hurwitz at each frozen time k , so that the matrix $A(k)$ has stable eigenvalues at each frozen time k . Let's note that p_0, \bar{p}, \bar{l} are function of β, θ_p , which are function of time k

$$\Delta p_0 = \frac{\Delta p_0}{\Delta \beta} \frac{\Delta \beta}{\Delta \theta_p} \Delta \theta_p, \quad \Delta \bar{p} = \frac{\Delta \bar{p}}{\Delta \beta} \frac{\Delta \beta}{\Delta \theta_p} \Delta \theta_p, \quad \Delta \bar{l} = \frac{\Delta \bar{l}}{\Delta \beta} \frac{\Delta \beta}{\Delta \theta_p} \Delta \theta_p \quad (13)$$

The first factors $\in l_\infty$ if C1 holds, the second factors $\in l_\infty$ if M1 holds, and $\Delta \theta_p \in l_2$ thanks to E3. We conclude that $\Delta p_0, \Delta \bar{p}, \Delta \bar{l} \in l_2$. In addition $p_0, \bar{p}, \bar{l} \in l_\infty$. This imply $\|\Delta A\| \in l_2$ and, using Theorem 3 in the Appendix, we conclude that the homogeneous part of (9) is exponentially stable (e.s.).

Step 3. From Theorem 2 of Appendix we have, considering (9)

$$\|(y_p)_k\|_{2\delta} \leq c \|(\epsilon m_s^2)_k\|_{2\delta} + c, \quad \|(u_p)_k\|_{2\delta} \leq c \|(\epsilon m_s^2)_k\|_{2\delta} + c \quad (14)$$

for some $\delta > \min\{\lambda_0^2, \delta_0\}$, where $\lambda_0 \in (0, 1)$ is the exponential convergence rate of the omogenous part of (9). We define the fictitious signal $m_f^2(k) \triangleq 1 + \phi^T(k)\phi(k) + \|(y_p)_{k-1}\|_{2\delta}^2 + \|(u_p)_{k-1}\|_{2\delta}^2$. Since $\delta > \delta_0$ we can verify that $|\phi|, |m_s| < m_f$. Besides using (14)

$$m_f^2 \leq c + c \|(\epsilon m_s^2)_{k-1}\|_{2\delta}^2 \leq c + c \|(\epsilon m_s m_f)_{k-1}\|_{2\delta}^2 \quad \forall k \geq 0$$

Using the Discrete-Time Bellman-Grownwall Lemma [6],

$$m_f^2 \leq c + c \sum_{i=0}^{k-1} \left(\prod_{i < j < k} (1 + \delta^{k-1-j} \epsilon^2 m_s^2(j)) \delta^{k-1-i} \epsilon^2 m_s^2(i) \right)$$

Using the fact that geometric mean is less than arithmetic mean, and $\epsilon m_s \in l_2 \cap l_\infty$, we obtain

$$\begin{aligned} m_f^2 &\leq c + c \sum_{i=0}^{k-1} \left(\delta^{k-1-i} \epsilon^2 m_s^2(i) \left(1 + \frac{c}{k-1-i} \right)^{k-1-i} \right) \\ &\leq c + c e^c \sum_{i=0}^{k-1} \delta^{k-1-i} \epsilon^2 m_s^2(i) \leq c \end{aligned} \quad (15)$$

where we used again the fact that $\epsilon m_s \in l_2 \cap l_\infty$. We conclude that $m_f \in l_\infty$ and $\epsilon m_s^2 \in l_2 \cap l_\infty$. Using the boundedness of m_f , we can establish the boundedness of all signals in the closed-loop plant: $\phi, u_p, y_p, e_1 \in l_\infty$.

Step 4. Manipulating the the control law and the normalized estimation error equations, and using the characteristic polynomial equation and the Discrete-time Swapping Lemmas 1 and 4 [5], we can arrive to

$$e_1 = \frac{\Lambda}{A^*} v, \quad v = \bar{L} Q_m \frac{1}{\bar{\Lambda}_q} (\epsilon m_s^2) - \bar{L}(r_1 - r_2) + r_3 - r_4 - r_5 - r_6 \quad (16)$$

where r_1, r_2 are defined as in Swapping Lemma 1 with $W = Q_m/\bar{\Lambda}_q$, r_3 is as defined in Swapping Lemma 4 with $f = Q_m/\Lambda u_p$ and r_4, r_5, r_6 are as defined in Swapping Lemma 4 with $f = e_1, \Lambda_0 = \Lambda$. Due to $u_p, y_p, e_1 \in l_\infty$ and $\Delta l, \Delta p, \Delta \theta_a, \Delta \theta_b \in l_2$, it follows that $r_1, r_2, r_3, r_4, r_5, r_6 \in l_2$. Since $\Lambda \bar{L} Q_m / (A^* \bar{\Lambda}_q)$ is e.s., we conclude that $e_1 \in l_2$. Besides, we can prove that Δe_1 is bounded, which implies that $e_1 \rightarrow 0$ as $t \rightarrow \infty$.

Case (2):

Step 1. Following the same steps of the ideal case, we can arrive to expressions (9)-(11). The modelling errors due to Δ_m, d do not appear explicitly in (9)-(11). The difference with respect to the ideal case is that

$$\epsilon, \epsilon m_s, \Delta \theta_p \in \mathcal{S}\left(\frac{\eta^2}{m_s^2}\right) \quad (17)$$

Step 2. Like the ideal case $A(k)$ has stable eigenvalues at each frozen time k . Besides, $\|\Delta A\| \in \mathcal{S}(\frac{\eta^2}{m_s^2})$. From the form of $\eta(k)$ it follows that

$$|\eta(k)| \leq \Delta_1 \|(u_p)_k\|_{2\delta_0} + \Delta_2 \quad (18)$$

with Δ_1, Δ_2 as defined in Theorem 1. Let's note that $|\eta(k)|/m_s \leq \Delta_1 + \Delta_2$: therefore, $\epsilon, \epsilon m_s, \Delta\theta_p, \Delta A \in \mathcal{S}(\mu^2)$, with $\mu^2 \triangleq c(\Delta_1^2 + \Delta_2^2)$. The conclusion is that $A(k)$ is e.s. provided that $c(\Delta_1^2 + \Delta_2^2) < \mu^*$ for some μ^* . This last condition may not be satisfied, even for small Δ_1 unless d_0 is small enough: one way to satisfy the condition is to opportunely design the filter $F_\eta(z)$. We continue supposing the condition satisfied, so that $\Phi(k_2, k_1) \leq \beta_0 \lambda_0^{k_2 - k_1}$ for some $\beta_0 > 0, 0 < \lambda_0 < 1$ and $k_2 \geq k_1 > 0$.

Step 3. As in the ideal case, let $\delta > \min\{\lambda_0^2, \delta_0\}$: we can show, using the Discrete-time Bellman-Gronwall Lemma and the fact that $\epsilon m_s \in \mathcal{S}(\mu^2) \cap l_\infty$, that for $\delta < 1/(1+\mu^2)$ we can again arrive to (15), and conclude that $m_f \in l_\infty$ and $\epsilon m_s^2 \in \mathcal{S}(\mu^2) \cap l_\infty$.

Step 4. A bound for the tracking error e_1 is obtained by expressing e_1 in terms of signals that are guaranteed by the adaptive law to be of the order of the modelling error in the mean square sense. The tracking error equation has exactly the same form as in the ideal case and is given by

$$e_1 = \frac{\Lambda(z)z^{n-1}Q_m(z)}{A^*(z, k)} \epsilon m_s^2 + \frac{\Lambda(z)\alpha_{n-2}^T}{A^*(z, k)} \nu_0 \quad (19)$$

where ν_0 is the output of proper stable transfer functions whose inputs are elements of $\Delta\theta_p$ multiplied by bounded signals. Because $\Delta\theta_p, \epsilon m_s^2 \in \mathcal{S}(\mu^2)$, it follows that $e_1 \in \mathcal{S}(\mu^2)$.

4 Conclusions

We have presented the adaptive mixing control approach in a discrete-time setting. A key contribution is the stability and analysis of the adaptive mixing scheme. For the nominal and noiseless case, it was shown that the adaptive mixing control scheme drives the plant states to zero. In the presence of unmodeled dynamics and bounded disturbances that satisfy specified bounds, it was shown that the closed-loop state remains bounded and the tracking error is of the order of the modeling error.

Appendix

Theorem 2. Consider the LTV system given by

$$\begin{aligned} x(k+1) &= A(k)x(k) + B(k)u(k), & x(0) &= x_0 \\ y(k) &= C^T(k)x(k) + D(k)u(k) \end{aligned} \quad (20)$$

If the state transition matrix $\Phi(k_2, k_1)$ of (20) satisfies

$$\|\Phi(k_2, k_1)\| \leq \beta_0 \lambda_0^{k_2 - k_1} \quad (21)$$

for some $\beta_0 > 0$, $0 < \lambda_0 < 1$, and $u \in l_{2e}$, then for any $\delta \in (\delta_1, 1]$, where $\delta_1 > \lambda_0^2$ is arbitrary, we have

$$\begin{aligned} (i) \quad & |x(k)| \leq \frac{c\beta_0}{\sqrt{\delta - \lambda_0^2}} \|u_k\|_{2\delta} + \epsilon_k \\ (ii) \quad & \|x_k\|_{2\delta} \leq \frac{c\beta_0}{\sqrt{\delta^{-1}(\delta - \delta_1)(\delta_1 - \lambda_0^2)}} \|u_k\|_{2\delta} + \epsilon_k \\ (iii) \quad & \|y_k\|_{2\delta} \leq c_0 \|u_k\|_{2\delta} + \epsilon_k \end{aligned}$$

where $c_0 = \frac{c\beta_0}{\sqrt{\delta^{-1}(\delta - \delta_1)(\delta_1 - \lambda_0^2)}} \sup_k \|C(k)\| + \sup_k \|D(k)\|$, $c = \sup_k \|B(k)\|$, and ϵ_k is a term exponentially decaying to zero due to $x_0 \neq 0$.

Theorem 3. We are interested in studying the stability of linear systems of the form

$$x(k+1) = A(k)x(k) \quad (22)$$

Let the elements of $A(k)$ in (22) be bounded functions of time and assume that $|\lambda_i(A(k))| \leq \sigma_s$, $\forall i, \forall k \geq 0$, where $0 \leq \sigma_s < 1$. If any one of the conditions

$$\begin{aligned} (a) \quad & \|\Delta A(k)\| \leq \mu \\ (b) \quad & \sum_{i=k}^{k+N-1} \|\Delta A(i)\|^2 \leq \mu^2 N + \alpha_0, \text{ that is } \|\Delta A(k)\| \in \mathcal{S}(\mu^2) \end{aligned}$$

is satisfied for some $\alpha_0, \mu \in [0, \infty)$ and $\forall k, N \geq 0$, then there exists a $\mu^* > 0$ such that if $\mu \in [0, \mu^*)$, the equilibrium state of (22) is uniformly asymptotically stable (u.a.s.) in the large (that is equivalent to e.s.). Besides if the following condition holds

$$(c) \quad \|\Delta A(k)\| \in l_2$$

the equilibrium state of (22) is uniformly asymptotically stable (u.a.s.) in the large.

References

1. M. Kuipers and P. A. Ioannou. Multiple model adaptive control with mixing. *Submitted to IEEE Transactions on Automatic Control*.
2. R. A. Nichols, R. T. Reichert, and W. J. Rugh. Gain scheduling for h-infinity controllers: A flight control example. *IEEE Trans. on Control System Technologies*, 1:69–79, 1993.
3. R. Reichert. Dynamic scheduling of modern-robust control autopilot designs for missiles. *IEEE Control System Magazine*, 12:35–42, 1992.
4. D. J. Stilwell and W. J. Rugh. Interpolation of observer state feedback controllers for gain scheduling. *IEEE Trans. on Automatic Control*, 44:1225–1229, 1997.
5. P. A. Ioannou and B. Fidan. *Adaptive Control Tutorial*. Advances in design and control, SIAM, 2006.
6. C.A. Desoer and M. Vidyasagar. *Feedback Systems: Input-Output Properties*. Academic Press, New York, 1975.

On geometry and scale of a stochastic chemostat

Marc Joannides¹ and Irène Larramendy-Valverde¹

IBM UMR 5149

Université de Montpellier 2, France

(e-mail: {Marc.Joannides,Irene.Larramendy-Valverde}@univ-montp2.fr)

Abstract. A chemostat is a fixed volume bioreactor in which microorganisms are grown in a continuously renewed liquid medium. We propose a stochastic model for the evolution of the concentrations in the single species and single substrate case. It is obtained as a diffusion approximation of a pure jump Markov process, whose increments are comparable in mean with the deterministic model. A specific time scale, related to the noise intensity, is considered for each source of variation. The geometric structure of the problem, usable by identification procedures, is preserved both in the drift and diffusion term. We study the properties of this model by numerical experiments.

Keywords: Chemostat, Diffusion approximation, Jump Markov process, Monte Carlo.

1 Introduction

1.1 The chemostat

The chemostat (chemical environment is static) is a laboratory device used to study the growth of microorganisms like yeast or bacteria. It consists in a growth chamber populated with one or more species in a liquid medium of fixed volume that is continuously renewed, see Figure 1.1. The inflow contains the nutrient used by the bacteria to grow and reproduce while the outflow eliminates the biomass together with the substrate. It was initially designed to measure the specific growth rate of the species under study, by maintaining constant the concentration in nutrient of the medium.

We consider here the single species B which uses the substrate S as its nutrient to grow and reproduce. The experimental conditions are determined by the substrate concentration in the influent S^{in} and by the dilution rate D . *Washout* occurs when the dilution rate is so fast that the increase of biomass within the growth chamber is not sufficient to balance the output. The quantities of interest chosen to characterize the system will be the concentrations of the substrate and of the biomass at each time. This state can only vary through the effect of the two mechanic actions (inflow and outflow) and the biological transformation, which is the increase of biomass by the uptake of substrate.

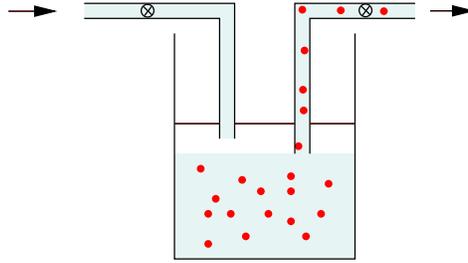


Fig. 1. Operating principle of a chemostat: the growth chamber is provided with a sterile liquid medium with a constant substrate concentration. The biomass is partly evacuated with the outflow

1.2 Deterministic model

The system is classically described by an ordinary differential equation based on a mass-balance principle (Bastin and Dochain[1]). We denote by b_t and s_t the respective concentrations of the biomass and of the substrate. Writing the balance for each component of the system yields the ODE

$$\dot{b}_t = \mu(s_t) b_t - D b_t \quad (1)$$

$$\dot{s}_t = -k \mu(s_t) b_t + D S^{\text{in}} - D s_t, \quad (2)$$

where $\mu(s_t)$ is the (bounded) *specific growth rate* of the species with limiting factor S and k is a stoichiometric coefficient. Numerous models have been proposed for the specific growth rate, among which the Monod model (uninhibited growth) and the Haldane model (inhibited growth) are the most commonly used. They read respectively

$$\mu(s) = \mu_{\max} \frac{s}{K_S + s} \quad \text{and} \quad \mu(s) = \mu_{\max} \frac{s}{K_S + s + \frac{s^2}{K_i}}$$

where the parameters μ_{\max} , K_s and K_i are to be estimated by statistical procedures. This dynamical system has been extensively studied with respect to many experimental conditions and growth models. To emphasize the geometric structure of the system, we write the vector form

$$\begin{pmatrix} \dot{b}_t \\ \dot{s}_t \end{pmatrix} = r(b_t, s_t) \begin{pmatrix} 1 \\ -k \end{pmatrix} + D \begin{pmatrix} 0 \\ S^{\text{in}} \end{pmatrix} - D \begin{pmatrix} b_t \\ s_t \end{pmatrix}. \quad (3)$$

with reaction kinetics $r(s_t, b_t) = \mu(s_t) b_t$, that exhibits the three vector fields corresponding to the three sources of variations.

To conclude this introduction, we notice that this deterministic description is based on the hypothesis that there are no stochastic fluctuations, or at least that they can be neglected. Moreover, it is also questionable whether

a space continuous formulation is still appropriate as far as washout is concerned. Indeed, the population of microorganisms in the apparatus could become so small that the concentration cannot any more be considered as varying continuously.

2 Pure jump model

Although the deterministic approach is widespread, there is a need to take in account the stochastic fluctuations inescapable when living organisms are involved. Various attempts to introduce a noise component in system (3) have been proposed, see e.g. Stephanopoulos *et al.*[12] or Imhof and Walcher[7]. This can be done by adding a diffusion coefficient to (3), or by adding a noise term to its discretized version. However, the *geometric structure* of the problem should be preserved by the perturbed system. Moreover, we should have in mind the discrete nature of the real state. For these reasons, we propose to modelize the phenomenon by a pure jump Markov process with three types of transitions associated with the three sources of variation. For each type of transitions, the mass balance should be true *in mean* and *at a given scale*. More precisely, let $(B_t, S_t)_{t \geq 0}$ be the pure jump Markov process whose transitions are described by

$$\mathbb{P}[(B_{t+h}, S_{t+h}) = (b', s') \mid (B_t, S_t) = (b, s)] = \begin{cases} h K^b r(b, s) + o(h) & \text{if } (b', s') = \frac{1}{K^b} (1, -k) \\ h K^{\text{in}} D + o(h) & \text{if } (b', s') = \frac{1}{K^{\text{in}}} (0, S^{\text{in}}) \\ h K^{\text{out}} D + o(h) & \text{if } (b', s') = \frac{1}{K^{\text{out}}} (-b, -s) \\ 1 - h [K^b r(b, s) + D (K^{\text{in}} + K^{\text{out}})] + o(h) & \text{if } (b', s') = (b, s) \\ 0 & \text{otherwise} \end{cases}$$

where K^b , K^{in} and K^{out} are scaling constant. This means that the process evolves only by jumps when an event E_Δ occurs, for $\Delta \in \{b, \text{in}, \text{out}\}$, corresponding respectively to the biological transformation, the inflow and the outflow. The rates $\lambda_\Delta(b, s)$ and the sizes $y_\Delta(b, s)$ of the jumps at a state (b, s) are given by the following table

| E_Δ | E_b | E_{in} | E_{out} |
|------------------------|-------------------------|--|-------------------------------------|
| $\lambda_\Delta(b, s)$ | $K^b r(b, s)$ | $K^{\text{in}} D$ | $K^{\text{out}} D$ |
| $y_\Delta(b, s)$ | $\frac{1}{K^b} (1, -k)$ | $\frac{1}{K^{\text{in}}} (0, S^{\text{in}})$ | $\frac{1}{K^{\text{out}}} (-b, -s)$ |

We also set to 0 the rates leading to a non admissible transition, that is a state outside the positive orthant.

2.1 Semimartingale representation

The generator of the process described above reads

$$Af(b, s) = \sum_{\Delta \in \{b, \text{in}, \text{out}\}} \lambda_{\Delta}(b, s) [f((b, s) + y_{\Delta}(b, s)) - f(b, s)], \quad (4)$$

for any f in its domain. For such an f , we have a semimartingale representation thanks to the Dynkin's formula

$$f(B_t, S_t) = f(B_0, S_0) + \int_0^t Af(B_s, S_s) ds + M_t^f$$

where M_t^f is a martingale. This formula remains valid for a wider class of function, even unbounded, provided some integrability condition holds, see e.g. Hamza and Klebaner [6] or Theorem 1.19 of Klebaner[8]. Since the specific growth rate is bounded, there exists $C > 0$ such that

$$\sum_{\Delta \in \{b, \text{in}, \text{out}\}} \lambda_{\Delta}(b, s) |y_{\Delta}(b, s)| \leq C(1 + |(b, s)|).$$

By Theorem 1.19 of Klebaner applied to the components of the identity function, we get the semimartingale representation for the process itself

$$\begin{aligned} B_t &= B_0 + \int_0^t [r(B_s, S_s) - D B_s] ds + M_t^B \\ S_t &= S_0 + \int_0^t [-k r(B_s, S_s) + D S^{\text{in}} - D S_s] ds + M_t^S \end{aligned}$$

which is the integral form of the SDE

$$\begin{pmatrix} dB_t \\ dS_t \end{pmatrix} = \left[r(B_t, S_t) \begin{pmatrix} 1 \\ -k \end{pmatrix} + D \begin{pmatrix} 0 \\ S^{\text{in}} \end{pmatrix} - D \begin{pmatrix} B_t \\ S_t \end{pmatrix} \right] dt + \begin{pmatrix} dM_t^B \\ dM_t^S \end{pmatrix}. \quad (5)$$

We see that the dynamics of our process is the sum of the drift appearing in (3) and of a martingale term carrying the stochastic perturbation.

3 Diffusion approximation

When the scaling parameters K^b , K^{in} and K^{out} are "reasonably" large, the process evolves by *small* but *frequent* jumps. In that case a diffusion approximation can be considered (Ethier and Kurtz[2]). Replacing the increments of f in (4) by a Taylor's expansion and dropping the terms of order greater than two gives, for $x = (b, s)$

$$\begin{aligned} \tilde{A}f(x) &:= \sum_{\Delta \in \{b, \text{in}, \text{out}\}} \lambda_{\Delta}(x) \left[\nabla f(x)^* \cdot y_{\Delta}(x) + \frac{1}{2} y_{\Delta}^*(x) \cdot H_f(x) \cdot y_{\Delta}(x) \right] \\ &= \nabla f(x)^* \cdot \sum_{\Delta \in \{b, \text{in}, \text{out}\}} \lambda_{\Delta}(x) y_{\Delta}(x) + \frac{1}{2} \sum_{\Delta \in \{b, \text{in}, \text{out}\}} \lambda_{\Delta}(x) y_{\Delta}^*(x) \cdot H_f(x) \cdot y_{\Delta}(x) \end{aligned}$$

where $H_f(x)$ denotes the Hessian matrix of f . \tilde{A} is the generator of a diffusion process $\tilde{X}_t = (\tilde{B}_t, \tilde{S}_t)$ which is solution of the SDE:

$$d\tilde{X}_t = \sum_{\Delta \in \{\text{b, in, out}\}} \lambda_{\Delta}(\tilde{X}_t) y_{\Delta}(\tilde{X}_t) dt + \sum_{\Delta \in \{\text{b, in, out}\}} \sqrt{\lambda_{\Delta}(\tilde{X}_t)} y_{\Delta}(\tilde{X}_t) dW_t^{\Delta}$$

with independent standard brownian motions W^{b} , W^{in} and W^{out} . Expanding the sums yields the vector form

$$\begin{aligned} \begin{pmatrix} d\tilde{B}_t \\ d\tilde{S}_t \end{pmatrix} &= \left[r(\tilde{B}_t, \tilde{S}_t) \begin{pmatrix} 1 \\ -k \end{pmatrix} + D \begin{pmatrix} 0 \\ S^{\text{in}} \end{pmatrix} - D \begin{pmatrix} \tilde{B}_t \\ \tilde{S}_t \end{pmatrix} \right] dt \\ + \sqrt{\frac{r(\tilde{B}_t, \tilde{S}_t)}{K^{\text{b}}}} \begin{pmatrix} 1 \\ -k \end{pmatrix} dW_t^{\text{b}} + \sqrt{\frac{D}{K^{\text{in}}}} \begin{pmatrix} 0 \\ S^{\text{in}} \end{pmatrix} dW_t^{\text{in}} + \sqrt{\frac{D}{K^{\text{out}}}} \begin{pmatrix} \tilde{B}_t \\ \tilde{S}_t \end{pmatrix} dW_t^{\text{out}}. \end{aligned} \tag{6}$$

This last form is much comparable with (3). The trajectories are continuous (almost surely) and the drift term is the same. Moreover the geometric structure is preserved by the diffusion term. Indeed, the stochastic perturbation in the diffusion term appears as a sum of three independent gaussian noises, each one acting along a vector field corresponding to a source of variation. Finally, the scaling parameters can be reinterpreted as noise intensity on the sources. Of course, it is possible to rewrite (6) as an SDE driven by a single two-dimensional brownian motion. However this would break the geometric understanding of the dynamics given by (6).

It should be noted that this diffusion model should be used away from the axis. In particular, studying extinction time would not make sense, see Pollett[10]. Finally, we observe that the diffusion coefficient vanishes as the scaling parameters tend to infinity, so that we get back the deterministic model.

4 Simulation algorithms

The pure jump model is classically simulated with the stochastic simulation algorithm, a.k.a Gillespie algorithm (Gillespie [4]), described below :

1. Initialization: let $(b, s) \leftarrow (b_0, s_0)$ and $t \leftarrow 0$
2. while $t < T_{\text{max}}$
 - compute global rate: $\lambda(b, s) = \sum_{\Delta \in \{\text{b, in, out}\}} \lambda_{\Delta}(b, s)$
 - compute next time event (exponential): $t \leftarrow t + \mathcal{E}(\lambda(b, s))$
 - jump: $(b, s) \leftarrow (b, s) + y_{\Delta}(b, s)$, where y_{Δ} is chosen with probability $\frac{\lambda_{\Delta}(b, s)}{\lambda(b, s)}$.

For large scaling parameters, all rates may be so fast that the procedure described above becomes unnecessarily slow. A number of variants have been proposed to speed up the procedure (review in Wilkinson[13]), mostly based on the approximation

$$(B_{t+h}, S_{t+h}) \simeq (B_t, S_t) + \sum_{\Delta \in \{\text{b, in, out}\}} N_{\Delta} y_{\Delta}(B_t, S_t)$$

where N_{Δ} denotes the number of event E_{Δ} that have occurred within $[t, t+h[$. We introduce

Assumption (i): The step size h is supposed to be small enough so that the rates λ_{Δ} do not vary significantly in the interval.

In that case, N_{Δ} are Poisson variables of respective parameters $\lambda_{\Delta}(B_t, S_t)h$. This leads to the *Poisson timestep method*, described by the algorithm:

1. Initialization: let $(b, s) \leftarrow (b_0, s_0)$ and $t \leftarrow 0$
2. while $t < T_{\max}$
 - For $\Delta \in \{\text{b, in, out}\}$, draw $N_{\Delta} \sim \mathcal{P}(\lambda_{\Delta}(b, s)h)$
 - Commit events: $(b, s) \leftarrow (b, s) + \sum_{\Delta \in \{\text{b, in, out}\}} N_{\Delta} y_{\Delta}(b, s)$
 - Increment time: $t \leftarrow t + h$

The timestep can also be adaptive, as in the *tau-leap method*, see Gillespie[5].

Still following Gillespie[3], we note that the diffusion process introduced above via a Taylor expansion, appears also as a numerical approximation of the jump process. Indeed, consider now

Assumption (ii): The timestep h is sufficiently large so that many events have occurred within $[t; t+h[$.

We can then use the normal approximation of the Poisson law, to get

$$\mathcal{P}(\lambda_{\Delta}(b, s)h) \simeq \lambda_{\Delta}(b, s)h + \sqrt{\lambda_{\Delta}(b, s)}\sqrt{h}\mathcal{N}^{\Delta}(0, 1)$$

and

$$\begin{aligned} (B_{t+h}, S_{t+h}) \simeq (B_t, S_t) + h \sum_{\Delta \in \{\text{b, in, out}\}} \lambda_{\Delta}(b, s) y_{\Delta}(b, s) \\ + \sqrt{h} \sum_{\Delta \in \{\text{b, in, out}\}} \sqrt{\lambda_{\Delta}(b, s)} \mathcal{N}^{\Delta}(0, 1) y_{\Delta}(b, s) \end{aligned}$$

which is nothing else but a Euler discretization scheme applied to the SDE (6). As a result, numerical solutions of $(\tilde{B}_t, \tilde{S}_t)$ obtained by such a scheme will have approximately same behaviour as (B_t, S_t) sampled with time step h . However, the choice of the time step remains problematic since it has to meet the two antagonist requirements (i) and (ii).

The numerical simulations presented below use a Monod model for the growth rate. Table 1 shows the values of the parameters. We use the Euler–Maruyama scheme to simulate the solutions of the SDE involved, see Kloeden and Platen[9].

| k | μ_{\max} | K_S | D | S^{in} | K^b | K^{in} | K^{out} |
|-----|--------------------|-------|-----------------------|-----------------|--------|-----------------|------------------|
| 10 | 3 h^{-1} | 6 g/l | 0.12 h^{-1} | 0.5 g/l | 10^7 | 10^5 | 10^5 |

Table 1. Parameter values

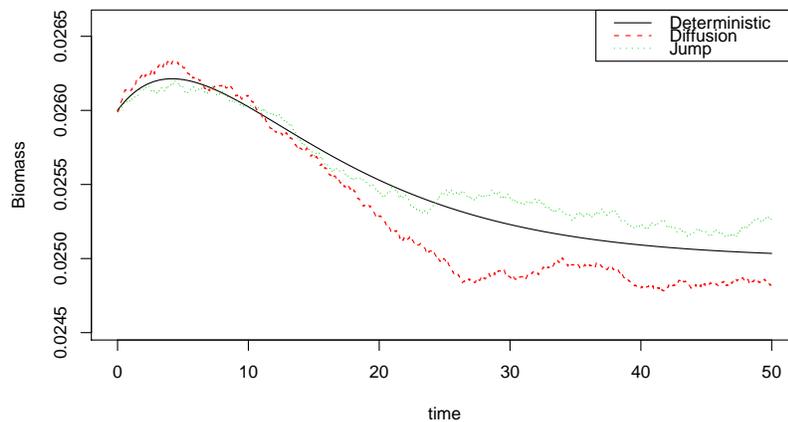


Fig. 2. Evolution of the biomass concentration for the deterministic model b_t , the diffusion-approximation \tilde{B}_t and the jump process B_t .

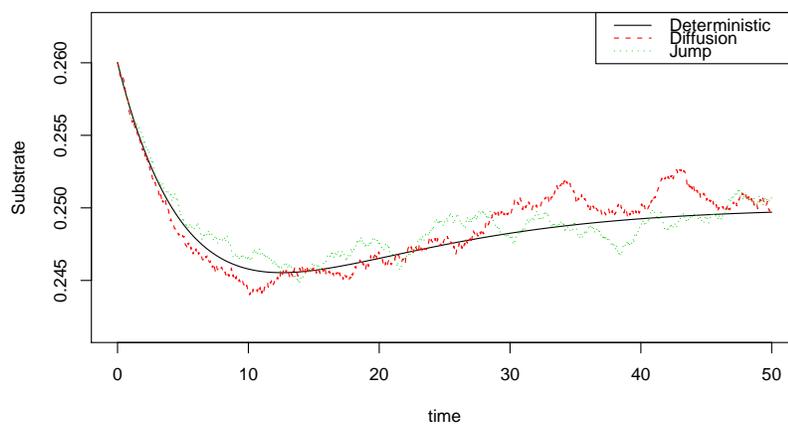


Fig. 3. Evolution of the substrate concentration for the deterministic model s_t , the diffusion-approximation \tilde{S}_t and the jump process S_t .

5 Conclusion

In this paper we have presented a way to account for randomness in a simple chemostat, while preserving the geometric structure. Stochasticity is first introduced with a pure jump Markov process whose infinitesimal increments agree with the classical deterministic model at a specific time scale. Using integrability conditions, we obtained the stochastic differential equation satisfied by this process.

Even if the fundamental structure of the system is discrete, it is reasonable to describe it by a process with continuous trajectories. We therefore introduced the diffusion approximation of the jump process, still preserving the geometry. The constants corresponding to the specific time scales can then be interpreted as the intensities of the independent noises affecting each source of variation. Numerical experiments showed that the approximation is safe provided that the system is far from "washout". It is expected that the geometry of the process will lead to efficient statistical procedures.

References

1. G. Bastin and D. Dochain. *On-line estimation and adaptive control of bioreactors*. Elsevier, Amsterdam, 1990.
2. Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes – Characterization and Convergence*. John Wiley & Sons, New York, 1986.
3. Dan T. Gillespie. The chemical langevin equation. *J. Chem. Phys.*, 113:297–306, 2000.
4. D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
5. F. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115:1716–1733, 2001.
6. K. Hamza and F. C. Klebaner. Conditions for integrability of markov chains. *Journal of Applied Probability*, 32(2):541–547, 1995.
7. L. Imhof and S. Walcher. Exclusion and persistence in deterministic and stochastic chemostat models. *Journal of Differential Equations*, 217(1):26–53, 2005.
8. Fima C. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, London, September 1998.
9. P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer Verlag, New York, 1992.
10. Philip Pollett. Diffusion approximations for ecological models. In *Proceedings of the International Congress of Modelling and Simulation*, 2001.
11. Hernán G. Solari and Mario A. Natiello. Stochastic population dynamics: The poisson approximation. *Phys. Rev. E*, 67(3):031918, Mar 2003.
12. G. Stephanopoulos, R. Aris, and A.G. Fredrickson. A stochastic analysis of the growth of competing microbial populations in a continuous biochemical reactor. *Mathematical Biosciences*, 45:99–135, 1979.
13. Darren J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, 2006.

The Sequential Feature of the Autoregressive Model

Christos P. Kitsos

Department of Mathematics, Technological Educational Institute of Athens, Egaleo, Athens, Greece

Email: xkitsos@teiath.gr

Abstract: The iterative schemes play a dominant role on evaluating parameters in Statistics. The target of this paper is to consider the first order Autoregressive Model, so important in Econometrics, as a “sequential approach” and construct the appropriate confidence intervals. A simulation study for different sample sizes provide evidence that the method performs rather well when the sample size is small. The normality is also investigated, in cases with small sample size.

Keywords: Sequential design, Optimal Design, Autoregressive Model

1 Introduction

The sequential procedures, especially when the optimal experimental design is adopted, arise the question if Fisher’s information matrix is valid, as the observations are not independent, Ford et. al. (1985). Moreover when the sequential design approach is adopted, two lines of thought are considered: a fully sequential design, or an equal batch sequential approach, Kitsos (1989). The fully sequential design has been defined as that one, where the number of observations, at each stage, coincide with the number of the parameters involved. Therefore for the Autoregressive model one observation is devoted at each stage.

From an Econometric point of view, see Pindyck and Rubinfeld (1988) among others, it is rather difficult to consider an experimentation adopting the model, but for the Statistical Signal Process point of view, it can be certainly accepted, as of practical use.

Therefore, in this paper, we face the Autoregressive model for both lines of thought, as we believe both are complementary.

Linear filter Input/Output (I/O) relations are well developed, adopting for the discrete time system, convolution representation of a linear system. Let X_t , $t \geq 1$ be a random discrete time random process with mean $\mu_t = EX_t$, and covariance

$$\text{Cov}(t, t') = E\{(X_t - \mu_t)(X_{t'} - \mu_{t'})\}.$$

Let h_k be the Kronecker δ -response of a discrete time linear filter. For the output process Y_t (described by the convolution integral) we consider

$$Y_t = \sum X_{t-k} \delta_k, \quad (1)$$

in the sense that, for an event ω , from the underlying probability space Ω , holds

$$Y_t(\omega) = \sum X_{t-k}(\omega) \delta_k, \quad (1a)$$

with probability one.

Now, consider that X_t is an autocorrelated discrete time two-sided random process, with mean μ and variance σ^2 with causal- δ -response

$$h_t = \theta^k, \quad k \geq 0, \quad |\theta| < 1. \quad (2)$$

The expected value of Y_t is (as Y_t is weakly stationary),

$$\mu_t = EY_t = \mu \sum h_k = \mu \frac{1}{1-\theta}, \quad t \in \mathbb{Z}. \quad (3)$$

The output variance is

$$\begin{aligned} \text{Var}Y_t &= \sigma^2 \sum_{t=0}^{\infty} h_k h_{t-k} = \sigma^2 \sum_{t=k}^{\infty} h_k h_{t-k} = \\ &= \sigma^2 \sum_{t=k}^{\infty} \theta^t \theta^{t-k} = \sigma^2 \theta^{-k} \sum_{t=k}^{\infty} (\theta^2)^t = \\ &= \sigma^2 \frac{\theta^k}{1-\theta^2}, \quad k = |k| \in \mathbb{N}. \end{aligned}$$

At $k = 0$,

$$\sigma_Y^2 = \frac{1}{1-\theta^2}. \quad (4)$$

Therefore, from (4) as $|\theta| \rightarrow 1$, the output variance grows with no bound, but with $|\theta| < 1$ the variance exists (and the process is weakly stationary). Therefore the existence of the variance is a problem, as well as its size, as “large” variance offers “small” information. From (1) and (2), we have that

$$\begin{aligned} Y_t - \theta Y_{t-1} &= \sum_{t=0}^{\infty} \theta^t X_{t-k} - \theta \sum_{t=0}^{\infty} \theta^t X_{t-1-k} = \\ &= X_t + \sum_{t=1}^{\infty} \theta^t X_{t-k} - \theta \sum_{t=0}^{\infty} \theta^t X_{t-1-k} = \\ &= X_t. \end{aligned}$$

Hence,

$$Y_t = \theta Y_{t-1} + X_t. \quad (5)$$

This presentation provides evidence that X_t process represents “the new added information” to Y_t . If we are restricted to independent identically distributed (iid) (and not just uncorrelated) the relation (5) is called the first-order autoregressive model for the model, in contrast to the introduced term ordinary convolution representation (also known as moving- average model).

Relation (5) is the link between signal processing, econometric model and general statistical linear models. Indeed, the econometric approach is the other one which still faces practical problems due to the first-order autoregressive, and provides Least Square Estimates for the “unknown parameter” θ based on the following line of thought.

Consider the first-order autoregressive model

$$Y_t = \theta Y_{t-1} + \varepsilon_t, \quad (6)$$

where the errors ε_t 's are identically independently distributed from the $N(0, \sigma^2)$, with $N(\cdot, \cdot)$ being the normal distribution as usually.

The conventional Least Square Estimate of θ , for $|\theta| < 1$, is

$$\hat{\theta}_n = \hat{\theta}_{LSE} = \frac{\sum_{i=1}^n y_i y_{i-1}}{\sum_{i=1}^n y_{i-1}^2}, \quad (7)$$

where a stochastic (initial value) y_0 , it is assumed to come from $N(0, \frac{\sigma^2}{1-\theta^2})$, $|\theta| < 1$ and is independent of ε_t 's. Recall (4): if the variance is “large” the constructed confidence intervals might be also “large” and therefore of no, practical, interest. Therefore we need a line of thought that might overpass these theoretical and practical problems, and we proceed as in the next section.

2 Fully sequentially designs

The distribution of $\hat{\theta}$ could be used to provide inference about θ , but unfortunately it is not known, but only in limit. However, when $|\theta|$ is not close to zero, the asymptotic distribution (normal) does not approximate well the true distribution in finite samples. Moreover we can consider (7) as a ratio estimate, which results problems on the existence of the moments. One of the excellent (dichotomous) theorem in Magnus (1986, Theorem 7) provides a simple criterion were the sth-moment of a general ratio estimate of the form (7) exists or not, and a practical result to this theoretical consideration can be that for the LSE (7) the moments exist up to and including the order $n - 2$.

With the above discussion it is clear that investigation is needed on how appropriate confidence intervals can be evaluated.

The experiment design approach has a reasoning to be considered the model (6) due to (5), and the line of thought described in the first part of section 1.

For the particular values of θ near to ± 1 the influence to the design, is essential and therefore to the process. This paper tries to answer these questions, adopting a theoretical line of thought, as well as a realistic one, through 1000 simulations, in this section. We are also considering if the nominal level influences the construction of the confidence interval.

Recall the autoregressive model and the LSE (7). Actually, if we set $X_t = Y_{t-1}$ the sample information can be evaluated as

$$I_n = \frac{1}{\sigma^2} \sum_{i=1}^n y_{i-1}^2 = \frac{1}{\sigma^2} J_n. \quad (8)$$

That certainly means: we have considered the observations being independent, although there are not, and Fisher's information measure remains the same, Kitsos

(1989), Ford et. al (1985). This result is due to statistical general linear model approach for model (6), see Seber (1977) among others. Moreover, as we are considering one observation at the end points, each time, we are ‘experimented’ at the interval $[0, x_{i-1}]$, and therefore a D-optimal design is constructed each time. In such a case the limiting design will be a D-optimal one, Kitsos (1989). This particular case, D-optimality, among the optimality criteria is the only one providing limiting results, and applied to construct an optimal design: choose as the next observation, that one, which minimizes the (generalized) variance. For the binary response problems, considering a quadratic response, Fornius (2008) constructed a sequential c-optimal design, but there is not a limiting result for it. For the particular Autoregressive model we can think in terms of the following well known result: we design, each time, at the end point, and therefore the linear model provides D-optimal estimates, each time. If the whole experimentation (or procedure) is D-optimal is another story and needs particular investigation. Fisher’s information I_n is a typical quantity to investigate if it is maximum, and therefore the inverse, the covariance minimum. But the D-optimal design for the autoregressive model is obtained if and only if we design at the end point.

Lai and Siegmund (1983) proven that

$$I_n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0,1),$$

i.e.

$$J_n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, \sigma^2).$$

Therefore, a $(1-a)\%$ confidence interval can be evaluated as

$$\hat{\theta}_n \mp t_{n-1; 1-\frac{\alpha}{2}} \left(\frac{\text{RSS}}{n-1} J_n^{-1} \right)^{1/2}, \quad (9)$$

with RSS being the Residual Sum of Squares, i.e.

$$\text{RSS} = \sum_{t=1}^n Y_t^2 - \hat{\theta}_n \sum_{t=1}^n Y_{t+1} Y_t.$$

The confidence interval evaluated by (9) it would be exact in case that in (6) Y_{t+1} are fixed in advance or selected independently of the other Y_t . A simulation study based on $N=1000$ experiments was performed, for different nominal levels, $\alpha = 0.05$, $\alpha = 0.2$, different sample sizes, $n=10, 5$ observations, with $y_0 = 0.0$ and . The skewness (s), kurtosis (κ), the Mean Square Error (MSE), the estimated coverage C and the average $\bar{\theta}$ were evaluated for different true θ , even beyond the interval $(-1,1)$. The results are performed in Table 1 and Table 2, the first line is referred to $n = 10$, the second to $n = 5$, and provide evidence, we believe that the theoretical approach it is not completely applicable to small size data sets.

3 Discussion

From the above Tables it is easily realized that the choice of nominal level has an influence on the results: In both case the evaluated coverage probabilities are

unsatisfactory as are less than the expected, but in Table 2 is even worse due to the choice of the nominal level. But certainly are not that bad, with nominal level 0.05 so that to provide any evidence that the confidence interval evaluation as in (9) should not be adopted. As far as the normality concerned the results are unsatisfactory when the parameter it in not within the interval $(-1,1)$. The Mean Square Error (MSE) is, as it was expected, larger when the sample size is reduced to $n = 5$, from $n = 10$. But still with such a small sample size the results are not disappointed. Therefore although the data was selected “sequentially”, i.e. not independently, the ratio estimates evaluated, provide evidence, that we can ignore the sequential nature of the design, see also Kitsos (1989), and evaluate the (parametric) Fisher’s information

Table 1. Simulation Study for the sequential model (4), $\alpha=0.05$

| θ | C | MSE | s | κ | $\bar{\theta}$ |
|----------|------|------|-------|----------|----------------|
| -1.5 | .919 | .03 | 3.88 | 21.07 | -1.44 |
| | .968 | .27 | 1.13 | 5.79 | -1.29 |
| -1.0 | .947 | 0.09 | 1.39 | 5.47 | -.84 |
| | .969 | 0.27 | 0.08 | 10.00 | -.82 |
| -0.5 | .970 | 0.09 | 0.63 | 3.27 | -0.42 |
| | .979 | 0.25 | -0.09 | 6.06 | -0.39 |
| 0.0 | .967 | 0.10 | -0.06 | 2.56 | 0.00 |
| | .970 | 0.25 | 0.22 | 4.92 | 0.02 |
| 0.5 | .965 | 0.10 | -0.58 | 3.01 | 0.41 |
| | .973 | 0.33 | -0.68 | 4.95 | 0.79 |
| 1.0 | .955 | 0.09 | -1.29 | 5.45 | 0.85 |
| | .975 | 0.61 | -0.14 | 4.98 | 0.38 |
| 1.5 | .968 | 0.03 | -3.96 | 22.21 | 1.45 |
| | .857 | 0.29 | -1.23 | 6.02 | 1.29 |

Table 2. Simulation Study for the sequential model (4), $\alpha = .20$

| θ | C | MSE | s | κ | $\bar{\theta}$ |
|----------|------|------|-------|----------|----------------|
| -1.5 | .773 | 0.04 | 4.78 | 31.35 | -1.36 |
| | .764 | 0.25 | 1.56 | 6.98 | -0.88 |
| -1.0 | .778 | 0.09 | 1.12 | 4.64 | -.85 |
| | .851 | 0.27 | 0.65 | 4.83 | -.80 |
| -0.5 | .868 | 0.08 | 0.61 | 3.43 | -0.42 |
| | .862 | 0.27 | -0.48 | 7.83 | -0.39 |
| 0.0 | .856 | 0.09 | -0.11 | 2.73 | 0.00 |
| | .904 | 0.24 | 0.09 | 5.50 | 0.00 |
| 0.5 | .839 | 0.09 | -0.65 | 3.37 | 0.41 |
| | .866 | 0.25 | -0.28 | 4.83 | 0.38 |
| 1.0 | .827 | 0.07 | -1.38 | 5.84 | 0.87 |
| | .838 | 0.27 | -0.51 | 4.84 | 0.81 |
| 1.5 | .770 | 0.03 | -3.28 | 19.39 | 1.44 |

References

1. Ford, I., Titterington, D. M. and Wu C. F. J., Inference and Sequential Design, *Biometrika*, **72**, 545-551 (1985).
2. Fornius, E. F., Optimal Design of Experiments for the Quadratic Logistic Model, *PhD thesis, University of Stockholm, Sweden*.
3. Kitsos, C. P., The Simple Linear Calibration Problem as an Optimal Experimental Design, *Communications. In Statistics – Theory and Methods*, **31**, 1167-1177 (2002).
4. Kitsos, C. P., Fully sequential procedures in Nonlinear Design problems. *Computational Statistics and Data Analysis*, Vol. 8, pg. 13-19 (1989)
5. Lee, J. and Lung, R. B., Revisiting Simple Linear Regression with Autocorrelated Errors, *Biometrika*, **91**, 240-245 (2004).
6. Lon, T. L. and Siegmund, D., Fixed Accuracy Estimation of an Autoregressive Parameter, *Ann. Stat.*, **11**, 478-485.
7. Magnus, J. R., The exact moments of a ratio of quadratic forms in normal variables, *Annals D'Economie et De Statistique*, **4**, 95-109.
8. Nankevris, J. C. and Savin, N. E., The exact Moments of the Autoregressive Model: Corrections and Extensions, *J. of Econometrics*, **37**, 381-388 (1989).
9. Park, R. E. and Mitchell, B. M., Estimating the Autocorrelated Error Model with Trended Data, *Journal of Econometrics*, **13**, 185-201 (1980).
10. Pindyck, R. S., Rubinfeld, D. L., *Econometric Models and Economic Forecasts*, McGraw-Hill (1980).
11. Roy, A. Falk, B. and Fuller, W. A., Testing for Trend in the Presence of Autoregressive Error, *J. of Amer. Stat. Association*, **99**, 1081-1091 (2004).
12. Seber, G. A. F., *Linear Regression Analysis*, John Wiley (1977).

Evaluation of Robust Algorithms Sequence Designed to Eliminate Outliners from Cloud of 3D Points Representing 3D Human Foot Image

Samuel Kosolapov

Signal and Image Processing Laboratory, ORT Braude College
Karmiel, Israel
Email: ksamuel@braude.ac.il

Abstract: In order to optimize operation of inexpensive non-laser 3D Structured Light Scanner adapted for fast Human Foot 3D Imaging, sequence of robust algorithms was designed, implemented and tested. During multi-frame image acquisition, sequence of Averaging and Median filters was used to reduce camera noise of different origin. A number of Robust Color Edge detectors used in the step of 3D calculations were evaluated and compared. The least amount of outliners was obtained in the algorithm evaluating Color Edge position by using scalar product in the RGB space after grey component elimination. Resulted cloud of 3D points was additionally processed by a sequence of logical operators and robust Median and Gaussian filters in order to eliminate 3D outliners. Processed cloud of 3D Points was overlapped onto 2D color image of the human foot (combined from a relevant frames of multi-frame image), thus creating true color presentation of the Human Foot. After visual inspection on the 2D monitor (for different view angles), created cloud of 3D points was converted to standard STL file and routed to 3D Printer for individual insole production. Accuracy and fitness of created insoles were evaluated and found accurate enough for chosen application.

Keywords: Image Processing, 3D Imaging, 3D Scanner, Insole, Point Cloud, 3D Point Cloud, Structured Light Technique, Outliners Elimination, Robust Algorithms, Median Filters

1 Introduction

According to many specialists in podiatric medicine, problems, such as pain in the feet, knees, ankles, back and neck may be solved by usage of simple devices such as foot devices (insoles) matched individually to the patient. Mass production of "typical insoles" is inexpensive, but not satisfactory enough for many patients. Individual insole production in podiatry clinics by using plaster/foam templates is expensive (requires large amount of working hours of the qualified specialists) and, in some extent, inconvenient (dust in surgery clinic during mold treatment).

In order to overcome the mentioned above problems, it was proposed to measure 3D shape of the human foot by using any appropriate 3D scanner (3D scanning is a simple, clean and fast process) and to produce individual insole by routing measured XYZ points to CNC (or desktop CNC, 3D printer) positioned remotely (*).

State-of-the-art laser 3D scanners [1], [2], [3] can provide high XYZ accuracy, but their design requires usage of high precision mechanical elements, which leads to slow scan and to forbiddingly high price. Even more serious problem is that laser scanners are considered as not well suited for usage in the clinic because of danger of eye damage due to uncontrolled laser ray reflections.

Non-laser structured light 3D scanners exhibit lower XYZ accuracy and have well-known ambiguity problem in the real-life situations [3]. Time-sharing utilized in some inexpensive 3D scanners [4] solves ambiguity problem, but, practically, time sharing leads to forbiddingly slow scan.

The newly designed compact structured light 3D scanner device [10] utilizes special color edge pattern. The pattern consists of a sequence of colored, white and black strips, creating unique and non-unique color edges. This multi-strip pattern solves ambiguity problem and enables fast scan with reasonable XYZ resolution.

2 Robust Image Acquisition

Acquisition software of the device [10] was designed to grab a set of at N_S series of 2D images; each series consisting of N_F frames (2D images). Color slide containing multi-strip pattern was mechanically shifted after specific series acquisition completion. Typical N_S value was set to 4. Values of N_F varied from 1 to 100.

In order to lower the price, inexpensive USB CMOS camera was used for images acquisition. Preliminary tests revealed that CMOS camera noise and electronic distortions were such significant that in order to enable reliable color strips detection N_F must be higher than 5. Practically, total number of frames was near 50, which (together with delays required for proper operation of mechanical parts of the scanner) results in total acquisition time about 2-3 seconds. On acquisition start beep sound is heard, signaling that patient is asked not to move his foot until next beep, signaling that acquisition is finished.

On the pre-process stage (immediately after acquisition completion), all $N_S * N_F$ frames (still stored in the acquisition buffers) were passed through Directional Median Filter (DMF) in order to eliminate salt & pepper camera noise and camera synchronization spikes. Direction of the filtration was normal to the direction of the strips edges. Optimal Half-Width of the DMF was found to be 3.

On the next step, all frames of the specific series were passed through combined directional Median-LPF filter (DMLPF). Parameters of DMLPF were empirically

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
International Conference, *Chania, Crete, Greece, 8 - 11 June 2010*

chosen to reliably exclude invalid frames resulted from camera malfunctions and to additionally filter-out camera noise of different origin without compromise XYZ resolution. Optimal Half-Width of the DMLPF was found to be between 2 and 4.

3 Images Decolorizing and Alignment

N_s pre-processed images were decolorized and aligned by using 4 fixed markers of pre-defined shapes. Resulted N_s aligned and decolorized images (ADI) appeared as if the camera was positioned normally to Z-plane from the fixed distance and as if the surface of human foot was absolutely white. Additionally, this procedure significantly reduced camera lens distortion (alignment accuracy was close to practical camera limit ~ 1 pixel) and uneven field illumination effects [2].

4 Robust Color Edge Detector

A number of different edge detectors [1], [2], [3] and color edge detectors [5], [6], [7] are known. Specifically to the described 3D scanner, a number of known color edge detectors adapted to the goals of current research were evaluated [8], [9], [11].

The best results (practically zero numbers of outliers) were obtained by using the following procedure:

To simplify the discussion, procedure of "Cyan-Magenta" color edge position detecting will be described here.

After alignment (described in (3)), direction normal to color strips orientation becomes column direction.

For each row and for each column of specific ADI two groups of pixels were organized: Left Color Strip (LCS) and Right Color Strip (RCS). Typically, color strip width (CSW) was set to 6 pixels. Possible colors of the strip were fixed as colors of color slide: {Black, Cyan, Magenta, Yellow, Gray}. For each {R,G,B} pixel of LCS and RCS the "closest" color was calculated by the following way:

```
IF (R < BLACK_LEVEL)
  AND
  (G < BLACK_LEVEL)
  AND
  (B < BLACK_LEVEL)
THEN Pixel is Black

IF ( ABS(R-B) < NOISE_LEVEL )
  AND
  ( ABS(R-G) < NOISE_LEVEL )
  AND
  ( ABS(B-G) < NOISE_LEVEL )
THEN Pixel is Gray
```

```
MagentaValue = ScalarProduct ( {R,G,B}, {255,0,255} )  
CyanValue    = ScalarProduct ( {R,G,B}, {0,255,255} )  
YellowValue  = ScalarProduct ( {R,G,B}, {255,255,0} )
```

(Scalar product was calculated in {RGB} space after Gray component elimination)

```
IF (MagentaValue > CyanValue) AND (MagentaValue > YellowValue)
```

```
THEN Pixel is Magenta
```

```
IF (CyanValue > MagentaValue) AND (CyanValue > YellowValue)
```

```
THEN Pixel is Cyan
```

```
IF (YellowValue > CyanValue) AND (YellowValue > MagentaValue)
```

```
THEN Pixel is Yellow
```

In case inside LCS at least (SCW-n) Cyan pixels were found, and inside RCS at least (SCW-n) Magenta pixels were found, the pixel between LCS and RCS was marked as "color edge candidate" (CEC). Typically, n was set to 1 (as 5 against 1 "vote"). For each CEC, robust average colors of LCS and RCS were recalculated and their scalar product (SP) was calculated. Column of CEC having maximal SP was marked as "Cyan-Magenta" Edge position {ColumnOfMax, row}. In case "Cyan-Magenta" CEC was not found for the specific row, edge position was marked as invalid by setting Edge position as {-1, row}.

This procedure was repeated for each color edge of the color slide.

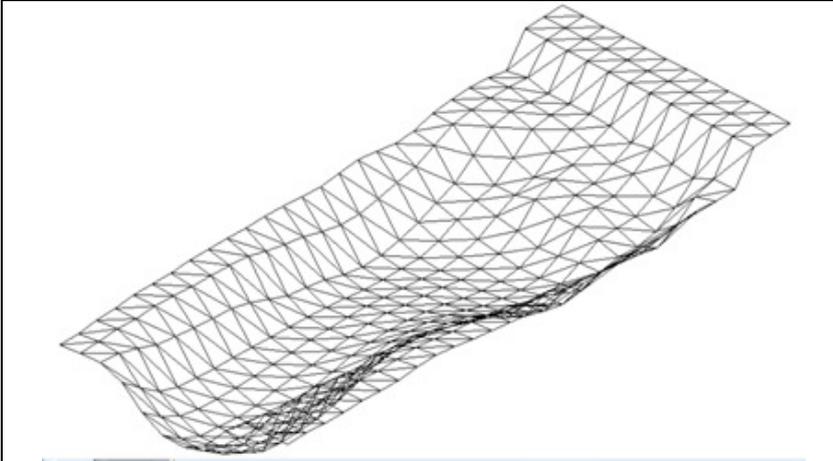


Fig. 1. CNC ready STL 3D Presentation of typical Human Foot

5 XYZ Cloud calculations and processing

Result of the color edge detector described in (4) is a plurality of curved lines. Every curved line was smoothed by using DMLPF (in the row direction). Then, by

using well known triangulation procedure [1], [3], cloud of XYZ points representing foot surface was generated.

XYZ points positioned outside predefined box were outlined. Resulted XYZ cloud was used to create a number of user friendly 2D and 3D presentations of human foot in test.

Additionally, standard STL file (see Fig. 1) was created and routed to 3D printer to evaluate accuracy of 3D Scanner. For the CMOS camera having VGA resolution 640x480, XYZ accuracy of the smoothed surface was evaluated as 0.5mm.

6 Conclusions

Inexpensive structured light 3D Foot Scanner enables creation of insole prototypes with accuracy adequate for the specified goal. Resulted XYZ cloud was effectively clean from dangerous for CNC /3D Printer outliners.

Acknowledgment:

* This study was supported by TNUFA Grant from the Chief Scientist of the Ministry of Trade and Industry, Israel, and by a grant from research committee of ORT Braude College, Israel.

References

1. Azad, P., Gockel, T., Dillmann R, Computer Vision. Principles and Practice, *Electron International Media BV* (2008).
2. Russ, J., The Image Processing Handbook. CRC Press (2007).
3. Jain. R., Kasturi, R., Schunck, B., Machine Vision. *McGraw-Hill Inc.*, (1995).
4. Danenberg, N., Kosolapov, S., Method and apparatus for determining the three dimensional shape and dimensions of the object. PCT Int. Appl. WO/2004/070315 (2004).
5. Androutsos, P., Androutsos, D., Plataniotis. K. N., Venetsanopoulos, A. N., Colour edge detectors: an overview and comparison. *IEEE 1997 Canadian Conference*. 607-610 vol. 2 (1997).
6. Tsang, W. H., Tsang, P. W. M., Edge gradient method on object color. *1996 IEEE TENCON. Digital Signal Processing Applications*. 304-310 vol. 1. (1996).
7. Trahanias, P. E., Venetsanopoulos, A. N., Color edge detection using vector order statistics. *Image Processing. IEEE Transaction*. 259-264 (1993).
8. Kosolapov, S., Comparison of accuracy of different structure light techniques. *in Applied Harmonic Analysis* (abstracts) (2007).

9. Kosolapov, S., Comparison of accuracy of different structure light techniques adapted for 3D human foot imaging. *Abstracts of the 3rd ORT Braude College Interdisciplinary Research Conference*. 31 (2007).
10. Kosolapov, S., Color edge based systems and method for determination of 3d surface technology. *US Patent. US 7,456,842 B2* (2008).
11. Kosolapov, S., Robust color edge detector for the compact 3D scanner. *Abstracts of the 4rd ORT Braude College Interdisciplinary Research Conference*. 46 (2008).

General parametric reliability model for repairable system

Makrem KRIT

Abstract

The aim of the paper is to expose a general form of modeling repairable system reliability. It is a bathtub form presented as a superposition of two Non-Homogeneous Poisson Processes (NHPP) and Homogeneous Poisson one (HPP). Moreover, the particularity of this model allows taking account of system state improvement in time course. The estimation of its parameters is considered through Maximum Likelihood (ML) and Expectation-Maximization (EM) algorithm. Decision tests are revealed to choose between a HPP and our model. Field failures data from an industrial setting are used to fit the model. In order to specify asymptotic properties, a Monte-Carlo simulation is employed, allowing to compare the estimate of our model by ML and EM algorithm. In this procedure, we are going to discuss two various cases of degradation.

Key words : repairable systems reliability, bathtub failure intensity, HPP, NHPP, estimation, likelihood, EM algorithm, Monte-Carlo simulation.

1 Introduction

Industrial world, that was made a long time, before do not cease to gain reliability and efficiency of their systems. The major stakes that can be placed in certainty are mainly safety, availability, costs and especially these of maintenance and lifetime. Near the industrial companies, we can recapitulate these stakes by the competitiveness and the safety which became a temptation responsible for the management of maintenance and for to improving the reliability objectives. It is in this direction that the methods of maintenance optimization by reliability (MOR) were developed, which optimize the maintenance programs on basis of the system functional analysis and the experience returns: the best maintenance adapted to the good site. Random models and statistical methods are used more and more to evaluate the industrial performance system in term of reliability.

However, separately the experience returns, when a maintenance program is chosen, we do not know its efficiency and its impact on the system operation. The objective is

thus to model the system lifespan and to quantify its degradation state or its failure, to appreciate the impact of a maintenance action on system behavior, and to find the actions in order to differ or to eliminate degradation, starting from the knowledge of events observation. In fact, the significant sorrow remains: at rest, there are the risks that can blame reliability, availability or the system safety. Particularly, a very significant characteristic to consider is the evaluation of the system failure intensity, and primarily the discovery at the appropriate time of its degradation. Moreover, to optimize the maintenance programs respecting the availability and to reduce the maintenance costs using the Maintenance Optimization by Reliability (MOR), as it was the case in Jiang-Ji-Xiao [11]. More clearly, it is a question on the one hand of building stochastic models of failures process and repairs of various systems, and on the other hand, of implementing the statistical methods to exploit the failures and maintenances data raised by experts with an aim to evaluate the performance of these systems.

Degradation concepts were often used to characterize the lifetime of systems, and to apply them to the maintenance action durations. In literature, several modelings of this appearance, the most known of which are the exponential law and its two principal alternatives; the Weibull model and the Gamma law, evoked by Ascher [1], Friedman-Gertsbakh [7]. In industrial context, the authors distinguished two fundamental types from data which is dependent in fact on two classes of systems; reparable and non-reparable ones. They presented subordinated probabilistic models, in particular the exponential law and the Homogeneous Poisson Processes (HPP), characterizing an absence of degradation, and which constitutes the base of reliability and maintenance modeling (see e.g, Ascher-Feingold [2], Cohen-Sacrowitz [5]).

From a more realistic point of view, and in order to appear the instant when the system degradation begins, two models were proposed: a simplest by Raftery-Ackman [16], who breaks up the failure rate on two levels; initially, it is equal to λ_1 up to one instant γ_0 , then it changes level beyond γ_0 to fix itself at a height λ_2 . The other, it was developed by Zacks [19], to study a formulation for which the failure rate is constant at the beginning then increases according to particular form as from the instant γ_0 . A more general formulation than the two last allowed models is the one interesting interpretation, evoked by Bertholon-Celeux [3]. So that he is able to study the degradation instant of a non-reparable system. It's lifespan which is modeled by means of a law which failure rate is selected in the following way; it is constant up to one instant γ_0 , which translates the absence of degradation until this date, then it increases according to general form of Weibull, translating therefore a state of degradation (Jiang-Ji-Xiao [9], Jiang-Murthy-Ji [10]).

By the same principle which is seen to specify the behavior change (i.e. appearance of degradation) of reparable system, Bertholon-Celeux [3] has a great contribution to model

the process of successive failures using Non Homogeneous Poisson Processes (NHPP) which intensity is $\lambda(t)$, having the same appearance as the failure rate for non-reparable system. The deferring Poisson process is thus a superposition of two processes: an HPP characterizing constant failure intensity equal to $\frac{1}{\eta_0}$, and a NHPP specifying the Weibull intensity, which starts from that γ_0 . In this case, the failure number until the instant t is distributed according to Poisson law with $\Lambda(t) = \int_0^t \lambda(u)du$ parameter. The Power law Process (PLP) proposed in Gianpaolo [8] as a special form of NHPPs, which is commonly used in the practical reliability analysis of complex repairable system. Lately the generalized exponential (GE) distribution, as a particular case of NHPPs was introduced in Rameshwar-Debasis [17] as much as an alternative to the Weibull law. The purpose of this article is to formulate a model, more general, more realistic, and aiming at the behavior evolution of reparable system during all its life.

2 Modeling of the system degradation

We move in this section about the failure intensity in bathtub form to formulating pace of such intensity on the three phases of the system life (see Mi [14]). Two forms are distinguished one from the other by a small change over the service life period. For the first form, as it indicates the hereafter figure, the failures process is modeled by superposition of three Poisson processes; the first and the third non-homogeneous and the second is homogeneous, of which the intensity is selected in following way :

It decline up to one instant noted by γ_0 , according to the function of the form $\frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right)$, translating the system improvement state in time course. After that, it's constant on a level $\frac{1}{\eta_0}$ (there will not be an advance of system degradation in this phase) up to an instant γ_1 which beyond the intensity increases in accordance with the form function $\frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2} \right)^{\beta_2-1}$, discovering a degradation case. This idea is originally proposed by Mudholkar-Srivastava [15] in the context of non-reparable system and in the context of complex system by Xie-Tang-Goh [18].

It to be proved that this degradation modeling comprises two terms (seconds in the two expressions of phase I and II) that one finds in Weibull process, the first for a shape parameter $\beta_1 < 1$ and the second for one $\beta_2 > 1$. Thus, the failure intensity is defined as the sum of three functions: a first is constant and equal to $\frac{1}{\eta_0}$, the second is equal to $\frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right)$ until the instant γ_0 , afterwards it's canceled, and the third function is null until the instant γ_1 later equalizes to $\frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2} \right)^{\beta_2-1}$. It's proceeded by admitting the assumption of perfect corrective maintenance (discover Lefebvre [12]), by the principle of competing risk already stated in Bertholon-Bousquet-Celeux [4] like an

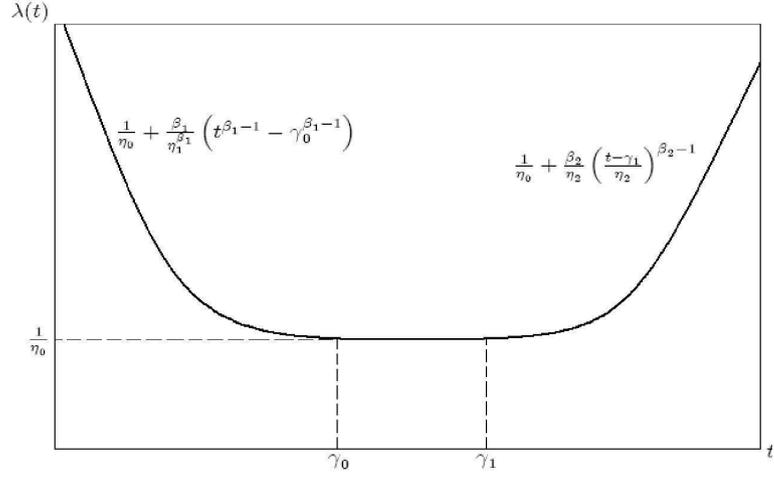


Figure 1: Graphic Modeling

alternative against Weibull law, the waiting duration of next failure can be written by the form $\mathcal{X} = \min(\mathcal{Y}, \mathcal{Z}, \mathcal{W})$, where:

- \mathcal{Y} a random variable, independent of \mathcal{Z} and \mathcal{W} , of Weibull law having as form the first expression, with a shift parameter equal to zero.
- \mathcal{Z} a random variable, independent of \mathcal{Y} and \mathcal{W} , of exponential law with parameter η_0 , which corresponds to constant failure intensity equalizes to $\frac{1}{\eta_0}$.
- \mathcal{W} a random variable of Weibull law having a shift parameter equal to γ_1 .

Our proposal, with the help of system behavior modeling, characterizes the failures process by intensity which is written as follows:

$$\lambda(t) = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2^{\beta_2}} \left(\frac{t-\gamma_1}{\eta_2} \right)^{\beta_2-1} & \text{if } t > \gamma_1 \end{cases} \quad (1)$$

Knowing this intensity, we can withdraw implicitly the system reliability, thanks to the following relation:

$$\mathcal{R}(t) = \exp \left\{ - \int_0^t \lambda(u) du \right\} \quad (2)$$

3 Parameters estimation of the model

3.1 Writing of likelihood function

The likelihood of data coming from a Poisson process, with general intensity $\lambda(t)$, is a function of seven parameters $\gamma_0, \gamma_1, \eta_0, \eta_1, \eta_2, \beta_1, \beta_2$, and which is written as follows:

$$\mathcal{L}(\theta; t_1, \dots, t_n) = \left[\prod_{i=1}^n \lambda_{t_i} \right] \exp \left\{ - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \lambda_s ds \right\} \quad (3)$$

In this case, intensity has not the same form front γ_0 , between γ_0 and γ_1 , and afterwards γ_1 . It's in this direction that requires distinguishing the γ_0 and γ_1 positions, in relation to n failure instants. With the aim of simplifying calculation, γ_0 and γ_1 will be often fixed in two failure instants t_i and t_j to hold in check $t_i < t_j$, particularly in degradation test treatment. For general case, where γ_0 and γ_1 are unspecified, that gives $\frac{n(n+1)}{2}$ possible likelihood forms, denoted $\mathcal{L}_{i,j}$ when one i observed failures front γ_0 and j failures between γ_0 and γ_1 . The likelihood function is developed in the same way by Bertholon-Bousquet-Celeux [4] is of the form:

$$\begin{aligned} \mathcal{L}_{i,j} = & \left[\prod_{k=1}^i \left(\frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right) \right] \times \left(\frac{1}{\eta_0} \right)^j \times \\ & \left[\prod_{z=j+1}^n \left(\frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t_z - \gamma_1}{\eta_2} \right)^{\beta_2-1} \right) \right] \times e^{-\left(\frac{\gamma_0}{\eta_1}\right)^{\beta_1} - \frac{1}{\eta_0} t_n - \left(\frac{t_n - \gamma_1}{\eta_2}\right)^{\beta_2}} \end{aligned} \quad (4)$$

with $1 \leq i \leq n$
 $i + 1 \leq j \leq n$

3.2 Property of maximum likelihood estimators (MLE)

Being fixed the instants γ_0 and γ_1 conditionally with failures data, the estimate program can be formulated thus as following:

$$\max_{\gamma_0, \gamma_1} \left(\max_{\eta_0, \eta_1, \eta_2, \beta_1, \beta_2} (\mathcal{L}_{i,j}(\eta_0, \eta_1, \eta_2, \beta_1, \beta_2)) \right)$$

The problem to pose is that likelihood is not limited. We can show the existence of a path in the parameters space which brings to infinitely increasing likelihood, when the shape parameters β_1 and β_2 themselves tend towards infinite one.

Demonstration (in appendix)

With the kept of preceding result, we arrive to distinguish β_1 and β_2 as two fixed and known parameters (for example $\beta_1 = \frac{1}{2}, \beta_2 = 2$), to be able to use the maximum likelihood procedure. Obviously, after testing the degradation existence unless he has sufficient information on system is allowed to know the dynamics of his degradation and to fix the two shape parameters.

3.3 The use of EM algorithm

The application of EM algorithm, proposed by Dempster-Laird-Rubin [6], in our model is on the one hand to discover an approximation of MLE in the case when γ_0 and γ_1 are fixed in observed failure values, $t_{(i)}$. On the other hand, we calculate, after that the likelihood obtained in each case corresponding to different (γ_0, γ_1) couples. Subsequently, we select the (γ_0, γ_1) case which leads to the greatest likelihood, the parameters values is obtained like the estimate of parameters true values. If the observations number is small, so much it is better to fix γ_0 and γ_1 elsewhere than in two observed failure values which are not sufficiently numerous.

4 Homogeneity test of failures process

We try in this paragraph to formalize procedure allowing to decide if the failures process is homogeneous or not. More precisely, in our context we prospect to know if there is one γ_0 to γ_1 period during which failures are purely accidental, and they trickle from the mechanism which integrates degradation elsewhere this period. The assumption test is thus defined as follows:

$$\begin{cases} H_0 : \text{the process is homogeneous Poisson} \\ H_1 : \text{the process is managed by our model} \end{cases}$$

A non-informative bayesian approach presentation of the test, such often the case in industrial field where the data are very few and much censored. It is a question within this framework to testing: $\begin{cases} H_0 : \beta_1 = \beta_2 = 1 \\ H_1 : \beta_1 \neq 1 \text{ and/or } \beta_2 \neq 1 \end{cases}$.

That study shows than the calculation of the bayesian factor \mathcal{B}_{NH-H} is complicated and difficult to interpret, because β_1 and β_2 parameters have an improper probabilistic laws. But this factor depends primarily on ratio made up of likelihoods integrated into η , of what sort : $\frac{\mathcal{L}_{NH}(\beta_1, \beta_2)}{\mathcal{L}_H}$, as in Martz-Waller [13]. The components of this ratio respectively represent relative likelihoods to Weibull process and HPP. In practice, for any sample \mathcal{T} , the function $\frac{\mathcal{L}_{NH}(\beta_1, \beta_2)}{\mathcal{L}_H}$ holds an unimodal form. It is equal to 1 in a first point $(\beta_1, \beta_2) = (1, 1)$, and in a second (β_1^0, β_2^0) . Geometrically and in space, the sample leading to rejection of the HPP is that for which volume (since to matter of double integrals) integrated in the square $([1, \beta_1^0] \times [1, \beta_2^0])$ and represent more than $(100 - \alpha)$ confidence of total volume. This volume is commented indeed like bayesian factor \mathcal{B}_{NH-H} .

Consequently, it appears interesting to clear up a decision rule directly starting from the ratio $\frac{\mathcal{L}_{NH}(\beta_1, \beta_2)}{\mathcal{L}_H}$. According to this decision rule, a sample involves rejection of HPP when:

$$\Pr \left(\frac{\mathcal{L}_{NH}(\beta_1, \beta_2)}{\mathcal{L}_H} > 1 / \mathcal{T} \right) \geq (100 - \alpha) \quad (5)$$

To define a critical region for our modeling, we first of all cover the discussion on data likelihood already stated by the equation (4). For this fact, we fix the parameters β_1 and β_2 without having infinite likelihood, and considering that the instants γ_0 and γ_1 coincide with two instants of observed value properly by $t_{(i)}$ and $t_{(j)}$. Obviously, the case $\gamma_0 = 0$ and $\gamma_1 = t_n$ corresponding to the HPP implies:

$$\mathcal{L}_{0,n} = \left(\frac{1}{\eta_0}\right)^n \times e^{-\frac{1}{\eta_0}t_n} \quad (6)$$

whence, we draw from it the following relation:

$$\begin{aligned} \mathcal{R}_{i,j}(\eta_0, \eta_1, \eta_2) &= \left[\prod_{k=1}^i \left(\frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - t_i^{\beta_1-1} \right) \right) \right] \\ &\times \left[\prod_{z=j+1}^n \left(1 + \beta_2 \frac{\eta_0}{\eta_2} \left(\frac{t_z - t_j}{\eta_2} \right)^{\beta_2-1} \right) \right] \times e^{-\left(\frac{t_i}{\eta_1}\right)^{\beta_1} - \left(\frac{t_n - t_j}{\eta_2}\right)^{\beta_2}} \end{aligned} \quad (7)$$

representing the likelihoods ratio : $\frac{\mathcal{L}_{i,j}}{\mathcal{L}_{0,n}}$.

The critical region making it possible to answer the above-mentioned assumption test of which decision variable equalizes with $\max \mathcal{R}_{i,j}(\eta_0, \eta_1, \eta_2)$ (see for example Kass-Raftery [11]), and it is defined in the following way:

$$\mathcal{W} = \{\max \mathcal{R}_{i,j}(\eta_0, \eta_1, \eta_2) \geq \mathcal{K}\} \quad (8)$$

The maximum is taken for i varying from 1 to n , and j varying from $i + 1$ to n , since we seek to compare $\mathcal{L}_{i,j}$ (for $i = 1, \dots, n$, and $j = i + 1, \dots, n$) with $\mathcal{L}_{0,n}$ which agrees on absence of improvement and degradation of system. A strong value of decision variable (rigorously higher than \mathcal{K}) thus plays in favor the alternative assumption. To carry out this test we could check the principal property of independence between the decision variable law and the parameter η_0 under the homogeneity assumption. In fact, we distinguish two cases one from the other according to check whether η_0 is known or not.

5 Numerical experiments

5.1 Application to real data

In a general gait, we initially propose to test the homogeneity of failures process while supposing η_0 as unknown, on basis of failure observed for reparable system. If the test makes allows showing existence from improvement and degradation, then we estimate the model parameters $\gamma_0, \gamma_1, \eta_0, \eta_1, \eta_2, \beta_1, \beta_2$, by means of EM algorithm. The direct

maximum likelihood method is used when one has sufficient information about system allowing to know its improvement and degradation characters in outline to fix β_1 and β_2 .

Whereas in the inexistence case of statistically significant improvement and degradation, the suitable model is the HPP. The inspection period is a phase of service life; the system is thus not detected by significant improvement and degradation.

The real data analysis is object to real example concerning reparable system (hydraulic pump) about nuclear sector of France which was used in Bertholon-Celeux [3]. The studied system retains a hydraulic pump on which we have the observation of 6 successive failures (18 months, 30, 82, 113, 121, 126).

The homogeneity test is first of all carried out, resulting that the decision variables equal to 1.766 and the critical probability associated, evaluated by achievements simulation of HPP, and takes the 25.13 value. The homogeneity assumption is then rejected. This decision is ensured with Kolmogorov-Smirnov test by critical probability evaluated to 0.63 ($KSSTAT = 0.667$). The alternative assumption is allowed and the failures process is governed by our model as the figure shows it hereafter.

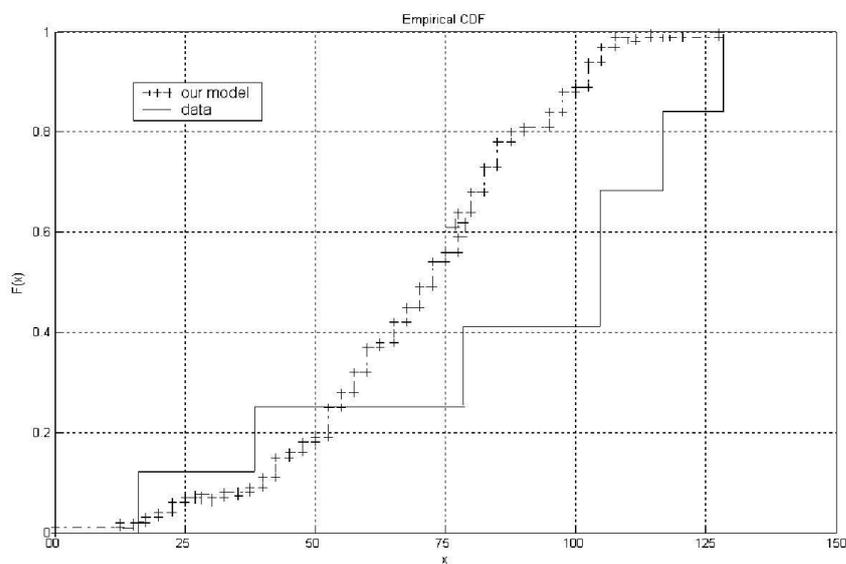


Figure 2: Empirical Cumulative Distribution Function

The estimate of model parameters using the EM algorithm gives the following results :

- γ_0 and γ_1 , respective instants of improvement end and degradation beginning are estimated to 26.685 and 101.412.
- The reverse of accidental failure rate η_0 is estimated by $\hat{\eta}_0 = 43.855$.

- The scale parameters η_1 and η_2 are estimated respectively by $\hat{\eta}_1 = 4.409$ and $\hat{\eta}_2 = 4.361$.
- The shape parameters β_1 and β_2 are estimated respectively by $\hat{\beta}_1 = 1.098$ and $\hat{\beta}_2 = 3.000$.

5.2 Simulation phase

In order to obtain concrete numerical results a Monte-Carlo simulation is employed, allowing to compare the estimate of our model by direct maximum likelihood (MLE) and by EM algorithm. We present two different cases:

- A first case retains 100 simulations of 50 size sample of our model with parameters $\eta_0 = 1, \eta_1 = 1, \beta_1 = 0.5, \eta_2 = 1, \beta_2 = 2, \gamma_0 = 30, \gamma_1 = 100$.
- A second case retains 100 simulations of 50 size sample of our model with the same parameters except for $\beta_2 = 3$.

The results are stated in form of mean and a 95% confidence interval. Things would be clearer studying the following table:

| | | First case ($\beta_2 = 2$) | | Second case ($\beta_2 = 3$) | |
|-----|------------------|------------------------------|--------------------|-------------------------------|--------------------|
| | | Mean | C I | Mean | C I |
| MLE | $\hat{\eta}_0$ | 1.705 | [1.411, 1.998] | 1.331 | [1.096, 1.565] |
| | $\hat{\eta}_1$ | 0.850 | [0.659, 1.041] | 0.968 | [0.810, 1.127] |
| | $\hat{\eta}_2$ | 0.880 | [0.656, 1.104] | 1.117 | [0.935, 1.299] |
| | $\hat{\gamma}_0$ | 30.454 | [25.130, 35.778] | 35.085 | [49.625, 60.546] |
| | $\hat{\gamma}_1$ | 104.723 | [99.153, 110.293] | 109.695 | [104.528, 114.863] |
| EM | $\hat{\eta}_0$ | 1.881 | [1.527, 2.234] | 1.156 | [0.985, 1.326] |
| | $\hat{\eta}_1$ | 0.801 | [0.592, 1.008] | 0.999 | [0.875, 1.123] |
| | $\hat{\beta}_1$ | 0.345 | [0.184, 0.506] | 0.454 | [0.411, 0.496] |
| | $\hat{\eta}_2$ | 0.799 | [0.559, 1.038] | 1.094 | [0.959, 1.230] |
| | $\hat{\beta}_2$ | 1.776 | [1.464, 2.089] | 2.783 | [2.602, 2.963] |
| | $\hat{\gamma}_0$ | 31.645 | [27.091, 36.198] | 32.560 | [27.493, 37.627] |
| | $\hat{\gamma}_1$ | 105.639 | [100.028, 111.250] | 109.229 | [104.459, 113.998] |

5.3 Discussion

Ultimately, subsequent the results of preceding tests, the failures process is a NHPP. The empirical cumulative distribution function of real data is evolved in the same direction as the simulated one. This process is then managed by our reliability model.

Hence, the effects of estimate go in front that there is improvement of system until the second failure (during 2.2 years of operation) and degradation starts from the fourth failure (beyond 8.5 years of operation). Considering the same unit of data over the improvement period and that of degradation, the scale parameters η_1 and η_2 over these two periods do not have a significant difference. This can be easily to ensure with skew of an averages difference traditional test.

The estimate value of β_2 is higher than 2, the failure intensity is increasing and convex announcing a marginal increase in degradation state. At the same time, β_1 takes an estimate value very near to 1 by saying that the intensity is practically constant. Thus, the failures are rather accidental and cannot be due to youth diseases. This purified model of improvement period, which is presented in Bertholon-Celeux [3] and Bertholon-Bousquet-Celeux [4], remains able alone to concretize the hydraulic pump behavior.

In light of simulations, we state the following criticisms :

- The $\hat{\eta}_j$ ($j = 0, 1$ ou 2) have the best behavior to one side for the first case where $\hat{\eta}_0$ appears to degrade.
- The $\hat{\gamma}_j$ ($j = 0$ ou 1) are all acceptable.
- The variability of $\hat{\beta}_j$ ($j = 1$ ou 2) is significant enough.
- In the aggregate, the EM procedures offer better estimators for the second case. The values of β_2 are rather higher than 2, then the curve is convex over the degradation period as it is presented by our model.

Nevertheless, a possible disadvantage of our model is that it implies seven parameters. In fact, it can be difficult to estimate these parameters for small-sized and/or censored samples. For this reason even, the MLE appear more reliable for industrial applications.

References

- [1] Ascher H. (1990). A survey of tests for exponentiality. *Communications in Statistics-Theory and Methods*, 19, 1811-1825.
- [2] Ascher H. and Feingold H. (1984). Repairable Systems Reliability. *New York and Basel, Marcel Dekker*.
- [3] Bertholon H. and Celeux G. (2001). Une modélisation du vieillissement. *Ph D. Thesis, Université Joseph Fourier, Grenoble, Fr.*

- [4] Bertholon H., Bousquet N. and Celeux G. (2004). An alternative competing risk model to the Weibull distribution in lifetime data analysis. *Technical Report* : RR-5265, INRIA Press, Orsay, France.
- [5] Cohen A. and Sacrowitz H.B. (1993). Evaluating Test for Increasing Intensity of a Poisson Process. *Technometrics*, 35, 446-458.
- [6] Dempster A.P., Laird N.M. and Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal. Roy. Statist. Soc. (Ser. B)*, 39, 1-38.
- [7] Friedman L. and Gertsbakh I. (1980). Maximum Likelihood Estimation in a Minimum-type Model with Exponential and Weibull Failure Modes. *J.A.S.A*, 75, 460-465.
- [8] Gianpaolo P. (2001). A Bounded Intensity Process for Reliability of Repairable Equipment. *Journal of Quality Technology*, 33, 4, 480-492.
- [9] Jiang R., Ji P. and Xiao X. (2003). Aging property of unimodel failure rate models. *Reliability Engineering and System Safety*, 79, 1, 113-116.
- [10] Jiang R., Murthy D.N.P. and Ji P. (2001). Models involving two inverse Weibull distributions. *Reliability Engineering and System Safety*, 73, 1, 73-81.
- [11] Kass R.E. and Raftery A. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773-795.
- [12] Lefebvre Y. (2003). Using the Phase Method to Model Degradation and Maintenance Efficiency. *International Journal of Reliability, Quality and Safety Engineering*, 10, 4, 383-405.
- [13] Martz H.F. and Waller R.A. (1982). Bayesian Reliability Analysis. *Wiley Series in Probability and Mathematical Statistics*, New york, 162-178.
- [14] Mi J. (1995). Bath-tub failure rate and upside-down bathtub mean residual life. *IEEE Transactions on Reliability*, 44, 388-391.
- [15] Mudholkar G.S. and Srivastava D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42, 99-102.
- [16] Raftery A.E. and Ackman V.E. (1986). Bayesian Analysis of a Poisson process with a Change-point. *Biometrika*. 73.1, 85-89.
- [17] Rameshwar D.G. and Debasis K. (2003). Discriminating between Weibull and Generalized Exponential Distributions. *Computational Statistics and Data Analysis*, 43, 179-196.

- [18] Xie M, Tang Y. and Goh T. N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety*, 76, 3, 279-290.
- [19] Zacks S. (1984). Estimating the Shift to Wear-out of Systems Having Exponential-Weibull Life Distributions. *Operations Research*, 32, 3, 741-749.

Appendix

The instants γ_0 and γ_1 are two fixed unspecified positive reals, but not necessarily in two respective instants t_i and t_j . The likelihood of the data can be written as :

$$\mathcal{L} = \left(\frac{1}{\eta_0}\right)^i \times \left[\prod_{k=1}^i \left(1 + \beta_1 \frac{\eta_0}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right) \right] \times \left(\frac{1}{\eta_0}\right)^j \times \left(\frac{1}{\eta_0}\right)^{n-i-j}$$

$$\left[\prod_{z=j+1}^n \left(1 + \beta_2 \frac{\eta_0}{\eta_2} \left(\frac{t_z - \gamma_1}{\eta_2} \right)^{\beta_2-1} \right) \right] \times e^{-\left(\frac{\gamma_0}{\eta_1}\right)^{\beta_1} - \frac{1}{\eta_0} t_n - \left(\frac{t_n - \gamma_1}{\eta_2}\right)^{\beta_2}}$$

where i and j represent the observations number of samples which are located respectively before the instant γ_0 and between γ_0 and γ_1 .

Thus we get the following log-likelihood function :

$$\ln \mathcal{L} = -n \ln \eta_0 + \left[\sum_{k=1}^i \ln \left(1 + \beta_1 \frac{\eta_0}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right) \right]$$

$$+ \left[\sum_{z=j+1}^n \ln \left(1 + \beta_2 \frac{\eta_0}{\eta_2} \left(\frac{t_z - \gamma_1}{\eta_2} \right)^{\beta_2-1} \right) \right] - \left(\frac{\gamma_0}{\eta_1}\right)^{\beta_1} - \frac{1}{\eta_0} t_n - \left(\frac{t_n - \gamma_1}{\eta_2}\right)^{\beta_2}$$

Let us fix in η_0 which can take an unspecified value, and we define the estimates of η_1 and η_2 parameters such as :

$$\tilde{\eta}_1 = t_i \text{ and } \tilde{\eta}_2 = t_n - \gamma_1.$$

At the point $(\eta_0, \tilde{\eta}_1, \tilde{\eta}_2, \beta_1, \beta_2)$, we obtain :

$$\ln \mathcal{L} = -n \ln \eta_0 + \ln \left(1 + \beta_1 \frac{\eta_0}{\tilde{\eta}_1^{\beta_1}} \left(t_i^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right)$$

$$+ \ln \left(1 + \beta_2 \frac{\eta_0}{\tilde{\eta}_2} \left(\frac{t_n - \gamma_1}{\tilde{\eta}_2} \right)^{\beta_2-1} \right) - \left(\frac{\gamma_0}{\tilde{\eta}_1}\right)^{\beta_1} - \frac{1}{\eta_0} t_n - 1$$

However, it proved implicitly as though :

$$1 + \beta_1 \frac{\eta_0}{\tilde{\eta}_1^{\beta_1}} \left(t_i^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) = 1 + \beta_1 \left[\frac{\eta_0}{t_i} - \frac{\eta_0}{t_i} \left(\frac{\gamma_0}{t_i} \right)^{\beta_1-1} \right] \longrightarrow -\infty \text{ when } \beta_1 \longrightarrow +\infty$$

in the same way $1 + \beta_2 \frac{\eta_0}{\tilde{\eta}_2} \left(\frac{t_n - \gamma_1}{\tilde{\eta}_2} \right)^{\beta_2-1} = 1 + \beta_2 \frac{\eta_0}{t_n - \gamma_1} \longrightarrow +\infty$ when $\beta_2 \longrightarrow +\infty$.
 $\ln \mathcal{L}$ is consequently infinite at the time or instant when $\beta_1 \rightarrow +\infty$ and/or $\beta_2 \rightarrow +\infty$.

Arithmetic Reduction of Intensity and Age models with bathtub failure intensity

Makrem KRIT * Abdelwaheb REBAÏ †

Abstract

In this paper, we will study the estimation of maintenance efficiency in Arithmetic Reduction of Intensity (ARI) and Arithmetic Reduction of Age (ARA) models with a memory m . These models have been proposed by Doyen (2005), the failure process is simply Non Homogeneous Poisson Process (NHPP). Our models are defined by reformulation of ARI and ARA ones using bathtub failure intensity. This form is presented like a superposition of two NHPP and Homogeneous Poisson one (HPP). Moreover, the particularity of this model allows taking account of system state improvement in time course. The maintenance effect is characterized by the change induced on the failure intensity before and after failure during degradation period. To simplify study, the asymptotic properties of failure process are derived. Then, the asymptotic normality of several maintenance efficiency estimators can be proved in the case where the failure process without maintenance is known. Practically, the coverage rate of the asymptotic confidence intervals issued from those estimators is studied.

Key words : repairable systems reliability, bathtub failure intensity, imperfect repair, maintenance, Homogeneous Poisson Process, Non Homogeneous Poisson Process, estimation, likelihood.

1 Introduction

However, separately the experience return, when a maintenance program is chosen, we do not know its efficiency and its impact on the system operation. The objective is thus to model the system lifespan and to quantify its degradation state or its failure, to appreciate the impact of a maintenance action on system behavior, and to find the actions in order to differ or to eliminate degradation, starting from the knowledge of events observation. Indeed, the significant sorrow remains: still there are the risks which can blame reliability, availability or

*Institut des Etudes Technologiques de Gafsa

†Ecole Supérieure de Commerce de Sfax

the system safety. Particularly, a very significant characteristic to consider is the evaluation of the system failure intensity, and primarily the discovery at the appropriate time of its degradation. Moreover, to optimize the maintenance programs respecting the availability and to reduce the maintenance costs using the maintenance optimization by reliability (MOR), as it was the case in Jiang-Ji-Xiao [11]. More clearly, it is a question on the one hand of building stochastic models of failures process and repairs of various systems, and on the other hand, of implementing the statistical methods to exploit the failures and maintenances data raised by experts with an aim to evaluate the performance of these systems.

The totality of momentous industrial systems is subjected to corrective and preventive maintenance actions which are supposed to prolong their functioning life. The efficiency evaluation of these maintenance actions is regarded with a great practical interest, but it was rarely studied. In literature, several maintenance effect models have been proposed, knowing for example, Pham-Wang [17] and Baxter-Kijima-Tortorella [3]. In their works, the authors try to index and to classify the various maintenance models. The majority of these models consider only the effect of CM; these models are known under the repair models name. These models are useful for modeling real systems which is sustained by constant repair. Several repair models, including the Brown-Proschan [5], Block-Borges-Savits [4], Kijima [14] models, the more general models of Dorado-Hollander-Sethuraman [7] and Last-Szekli [16], have all been useful in this regard. The Dorado-Hollander-Sethuraman model is much more general than the virtual age, reaching to take into account a great number of varied maintenance effects. Several theoretical properties of such models, including frequently estimators of the fundamental failure intensity and asymptotic confidence intervals, these estimators have been studied without assessing the maintenance efficiency. The same assumptions for these models can be also used for the PM models only.

The degradation concept summer is often employed to characterize the system's lifetime, and to apply them to the maintenance action durations. In the literature, several modelings of this appearance, of which most known are the exponential law and its two principal alternatives Weibull model and Gamma law, evoked by Ascher [1] and Friedman-Gertsbakh [12]. In industrial context, the authors distinguished two fundamental types from data and which are dependent in fact on two classes of systems; reparable and non-reparable ones. They presented subordinated probabilistic models, in particular exponential law and HPP, which characterize an absence of degradation, and which constitute the base of reliability and maintenance modeling (see e.g. Ascher-Feingold [2], Cohen-Sacrowitz [6]). The idea of our study is to develop some maintenance efficient estimators when the failure intensity in bath-tub form is known. Thus, the system behavior without maintenance is known, and the failure intensity is then supposed to be as a function of the single efficiency parameter ρ . For

this fact, we try to proceed in the same way as Doyen [8], by introducing in the first place the properties of maximum likelihood estimator, and in the second place by interesting in exposing an explicit estimator. Several works was carried out on the parametric statistical inference in imperfect repair models. We refer for example to the Shin-Lim-Lie [18] study in which authors developed a preventive maintenance policy; it is the same work for Yun-Choung [19]. For the case of *ARA* and *ARI* models, we evoke the Doyen-Gaudoin [10] and Doyen [9] works. The numerical results for our study were at the estimate base by the maximization likelihood method and its properties. Thus, it was necessary to initially study the behavior of failure process. We are interested in the case of the *ARI_m* and *ARA_m* reformulated models, like the most general case of these models classes.

2 Behavior of failure process

Our reformulation of the *ARI_m* and *ARA_m* models using failure intensity in bath-tub form is made to consider that the maintenance effect relates to several preceding failure instants. Obviously, during degradation period in view of the maintenance actions during improvement and service life two periods of system are supposed ABAO (i.e. they are carried out just to restore the system operation, considering that it's still without degradation). We thus build reformulations of the *ARI* and *ARA* models with memory m (*ARI_m* and *ARA_m*) by supposing that this effect relates to previous m failure instants occurring all lasting degradation period. Memory m represents the maximum number of previous instants preceding failure instants occurring in degradation phase which can influence the failure intensity, always reflects markovian property. The two *ARI_m* and *ARA_m* reformulated models are characterized by two failure intensities defined respectively by equations (1) and (2).

$$\lambda_t = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2} \right)^{\beta_2-1} - \rho \sum_{k=N_{\gamma_1}}^{\min(m-1, N_t-1)} (1-\rho)^k \frac{\beta_2}{\eta_2} \left(\frac{T_{N_t-k}-\gamma_1}{\eta_2} \right)^{\beta_2-1} & \text{if } t > \gamma_1 \end{cases} \quad (1)$$

and,

$$\lambda_t = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1-\rho \sum_{k=N_{\gamma_1}}^{\min(m-1, N_t)} (1-\rho)^k T_{N_t-k}}{\eta_2} \right)^{\beta_2-1} & \text{if } t > \gamma_1 \end{cases} \quad (2)$$

It is noticed that for the ARI_m model reformulation case, the maintenance effect does not make vary the slope of failure intensity, it continues to move similarly as before failure. When maintenance action is AGAN (of course beyond the instant γ_1), the slope of failure intensity is the same one as that of the renewed system (and not new in our model). For this reason, beyond the instant γ_1 , maintenance action cannot in any case be perfect. In this case, the efficiency parameter ρ cannot be higher than 1, allowing for intensity not to become negative. Then it is the case of our reformulation of the ARA_m model, correspondent to maintenance actions which are more than perfect.

Like any ARA model case of simple failure intensity, from the instant γ_1 the ARA_m reformulated models have failure intensity horizontally parallel to increasing part of initial intensity. In the same way, the ARI_m reformulated models have failure intensity vertically parallel to increasing part of the initial intensity.

We request to confirm the interesting property of the ARI_m and ARA_m models presented by Doyen-Gaudoin [10]. These models hold is what we call minimal and maximal degradation intensities. The minimal degradation intensity is concreted as a maximal lower limit for failure intensity. For our reformulation of the ARI_m model, this intensity is defined, as follows:

$$\text{For all } t > \gamma_1, \lambda_{\min}(t) = \begin{cases} (1 - \rho)^m \lambda(t) & \text{if } 0 \leq \rho \leq 1 \\ \lambda(t) & \text{if } \rho \leq 0 \end{cases} \quad (3)$$

Demonstration :

Consent to the failure intensity is increasing for all t of interval $]\gamma_1, +\infty[$, $\forall \mathcal{N}_{\gamma_1} \leq k \leq \mathcal{N}_t - 1$, $\lambda(\mathcal{T}_{\mathcal{N}_t - k}) \leq \lambda(t)$,

Consequently,

$$\lambda_{t-\gamma_1} \geq \lambda(t) - \rho \sum_{k=\mathcal{N}_{\gamma_1}}^{\min(m-1, \mathcal{N}_t-1)} (1 - \rho)^k \lambda(t) \quad (4)$$

$$\text{with } \lambda(t) = \frac{\beta_2}{\eta_2} \left(\frac{\mathcal{T}_{\mathcal{N}_t - k}}{\eta_2} \right)^{\beta_2 - 1}$$

In the same way,

$$\lambda_t \geq \lambda(t) \left[1 - \rho \sum_{k=\mathcal{N}_{\gamma_1}}^{m-1} (1 - \rho)^k \right] = (1 - \rho)^m \lambda(t)$$

In this fact :

$$\Pr(\lambda_t \geq (1 - \rho)^m \lambda(t)) = 1$$

Under the same condition, minimal degradation intensity of the ARA_m reformulated model is:

$$\lambda_{\min}(t) = \begin{cases} \lambda((1 - \rho)^m(t)) & \text{if } 0 \leq \rho \leq 1 \\ \lambda(t) & \text{if } \rho \leq 0 \end{cases} \quad (5)$$

By identical principle of minimal degradation intensity, we can similarly define maximal degradation intensity: it is expressed as a minimal upper limit for failure intensity. For our reformulations of the ARI_m and ARA_m models, maximal degradation intensity can be respectively defined by the two following relations:

$$\text{For all } t > \gamma_1, \lambda_{\max}(t) = \begin{cases} \lambda(t) & \text{if } 0 \leq \rho \leq 1 \\ (1 - \rho)^m \lambda(t) & \text{if } \rho \leq 0 \end{cases} \quad (6)$$

And,

$$\text{and, for all } t > \gamma_1, \lambda_{\max}(t) = \begin{cases} \lambda(t) & \text{if } 0 \leq \rho \leq 1 \\ \lambda((1 - \rho)^m(t)) & \text{if } \rho \leq 0 \end{cases} \quad (7)$$

The demonstrations of these last degradation intensities are very similar to that of the ARI_m reformulated model.

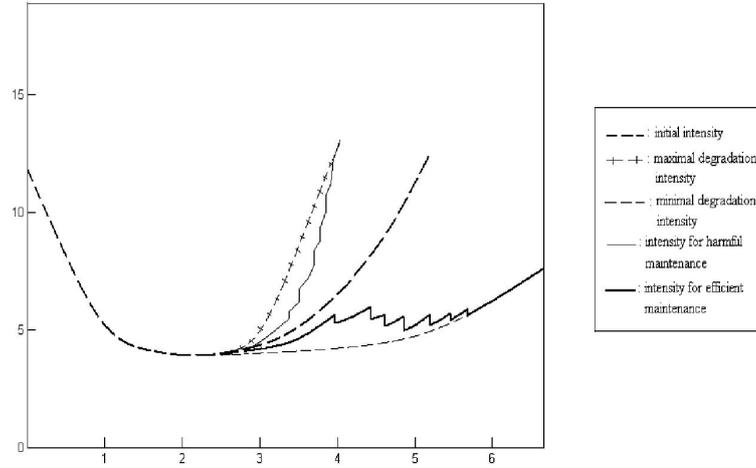


Figure 1: Failure intensity of ARA_3 reformulated model in harmful and effective maintenance cases.

The figure 1 represents the failure intensities of the ARI model with memory 3 like particular ARI_m case, for bath-tub failure intensity is defined by the parameters as: $\eta_0 = 26$, $\eta_1 = \eta_2 = 200$, $\beta_1 = 0.5$, $\beta_2 = 3$, $\gamma_0 = 1.6$, $\gamma_1 = 2.7$. These two failure intensities are defined one for effective maintenance ($0 < \rho < 1$) and the other for harmful maintenance ($\rho < 0$). We note in this representation for effective maintenance case, that in time course the difference between minimal degradation intensity and failure intensity tends to decrease. It's the same in harmful maintenance case, for the maximal degradation intensity.

Therefore, it appears in the effective maintenance case that the failure intensity and the minimal degradation intensity remain about equivalent. For the harmful maintenance case, the same property seems to be preserved but with the maximal degradation intensity. Indeed, we develop the asymptotic intensity (same principle which followed by Doyen [8]) like the minimal degradation intensity for the effective maintenance and the maximal degradation intensity for the harmful maintenance. Consequently, we admit for our reformulation of the ARA_m model, like asymptotic failure intensity, the function: $\lambda_\infty = \lambda((1 - \rho)^m(t))$. And the function defined by $\lambda_\infty = (1 - \rho)^m \lambda(t)$ is considered as an asymptotic failure intensity of the ARI_m reformulated model.

In continuation, the initial intensity, $\lambda(t)$, is supposed to be as a deterministic function, which is not identically null and which is increasing during the period of the system degradation (defined in Krit-Rebaï-Benbacha [15] without maintenance process). These conditions necessarily imply : $\lim_{t \rightarrow +\infty} \Lambda(t) = +\infty$. The function Λ is the cumulative failure intensity.

2.1 Asymptotic Behavior of failure process

In this paragraph, the idea is to show that the failure intensity (in particular it's increasing phase) and the asymptotic intensity have an identical behavior. Thus, we recall the property presented in Doyen [8], that if exists a function λ_{\min} , not decreasing and verifies for our model $\forall t > \gamma_1$: $\lambda_{\min}(t)$, hence for all $k \geq 0$: $t - \mathcal{T}_{N_t-k} = o(t)$

If moreover λ is a regular variation function, then for $t > \gamma_1$: $\lambda(t) - \lambda(t + o(1)) = o(\lambda(t))$.

Thereafter, the whole of asymptotic results of this study are rested on a rewriting of the failure intensity, considered exclusively by finished memory models. By means of $\rho \sum_{k=0}^{m-1} (1 - \rho)^k = 1 - (1 - \rho)^m$, this new form of failure intensity is defined, for the ARI_m reformulated model, $\forall t \geq \mathcal{T}_m \geq \mathcal{T}_{\gamma_1}$, as:

$$\lambda_t = \lambda_\infty(t) - \frac{\rho}{(1 - \rho)^m} \sum_{k=0}^{m-1} (1 - \rho)^k [\lambda_\infty(t) - \lambda_\infty(t + (t - \mathcal{T}_{N_t-k}))] \quad (8)$$

By means of the foregoing property, this formula is written as follows:

$$\begin{aligned} \lambda_t &= \lambda_\infty(t) - \frac{\rho}{(1 - \rho)^m} \sum_{k=0}^{m-1} (1 - \rho)^k [\lambda_\infty(t) - \lambda_\infty(t + o(1))] \\ &= \lambda_\infty(t) + o(\lambda_\infty(t)) \end{aligned}$$

In same way, for the ARA_m reformulated models, for all $t \geq \mathcal{T}_m \geq \mathcal{T}_{\gamma_1}$:

$$\lambda_t = \lambda_\infty(t) - \left[\lambda_\infty(t) - \lambda_\infty \left(t + \frac{\rho}{(1 - \rho)^m} \sum_{k=0}^{m-1} (1 - \rho)^k (t - \mathcal{T}_{N_t-k}) \right) \right] \quad (9)$$

And that :

$$\begin{aligned}\lambda_t &= \lambda_\infty(t) - [\lambda_\infty(t) - \lambda_\infty(t + o(1))] \\ &= \lambda_\infty(t) + o(\lambda_\infty(t))\end{aligned}$$

Consequently, for our reformulations of the ARI_m and ARA_m models, the failure intensity, for all $t > \gamma_1$, verify: $\lambda_t = \lambda_\infty(t) + o(\lambda_\infty(t))$. Under the same conditions, the cumulative failure intensity proves : $\Lambda_t = \Lambda_\infty(t) + o(\Lambda_\infty(t))$. This first order of asymptotic expansion of the failure intensity, make possible to verifies that the increasing phase of the failure intensity and the asymptotic intensity of the ARI_m and ARA_m reformulated models of finished memory have a same asymptotic behavior (like it was the remark on figure 1). In general case, this result is valid only if:

- The maintenance actions are not perfect: $\rho < 1$;

- The initial intensity of the ARA_m models is an increasing power function: $\lambda(t) = \alpha\beta t^{\beta-1}$, with $\alpha > 0$, $\beta > 1$;

These two properties are already checked in our study, $\rho < 1$ (the maintenance actions are not perfect) and the initial intensity has as an like increasing phase the function: $\frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2}\right)^{\beta_2-1}$. It is a well an increasing power function when $\eta_2 > 0$, $\beta_2 > 1$.

-The initial intensity of the ARI_m models is a function with increasing regular variations (see, for example, Embrechts-Klüppelberg-Mikosch [11]), which verify : There exists $\theta > 0$, for all $x > 0$, $\lim_{t \rightarrow +\infty} \frac{\lambda(xt)}{\lambda(t)} = x^\theta$.

This assumption is also checked in our model case, seeing that the increasing phase of the initial failure intensity verifies:

$$\exists \beta_2 > 1, \text{ for all } x > 0, \lim_{t \rightarrow +\infty} \frac{\frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{xt-\gamma_1}{\eta_2}\right)^{\beta_2-1}}{\frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2}\right)^{\beta_2-1}} = \lim_{t \rightarrow +\infty} \frac{(xt)^{\beta_2-1}}{t^{\beta_2-1}} = x^{\beta_2-1} \quad (10)$$

Using the second order of asymptotic expansion of the cumulative failure intensity, Doyen [8] goes more and expresses the difference between failure and asymptotic intensities. The author proved that the cumulative failure intensity for the ARI_m and ARA_m models with the power failure intensity. In consequence, for our reformulation with bath-tub failure intensity, the cumulative failure intensity of the ARI_m model can be written, for all $t \geq \mathcal{T}_m$, as:

$$\Lambda_t = \Lambda_\infty(t) + \frac{\rho}{(1-\rho)^m} \sum_{k=0}^{m-1} (1-\rho)^k \int_{\gamma_1}^t \lambda_\infty(s) - \lambda_\infty(\mathcal{T}_{\mathcal{N}_s-k}) ds \quad (11)$$

Thereafter, let's suppose that through the asymptotic intensity, or in an equivalent way, during the degradation phase of the system that the initial intensity is divergent. That's to say then the proposal that the cumulative failure intensity of the ARI_m reformulated models ensures:

$$\text{for all } t > \gamma_1, \Lambda_t = \Lambda_\infty(t) + \frac{1 - (1 + m\rho)(1 - \rho)^m}{\rho(1 - \rho)^m} \ln \lambda(t) + o(\ln \lambda(t)) \quad (12)$$

By analogy with the ARI_m reformulated models, and considering that the restriction of our initial intensity on $] \gamma_1, +\infty[$, is defined by:

$$\lambda_{] \gamma_1, +\infty[}(t) = \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t - \gamma_1}{\eta_2} \right)^{\beta_2 - 1}$$

then the cumulative failure intensity of the ARA_m reformulated models verify, for all $t > \gamma_1$:

$$\Lambda_t = \Lambda_\infty(t) + (\beta_2 - 1) \frac{1 - (1 + m\rho)(1 - \rho)^m}{\rho(1 - \rho)^m} \ln(t) + o(\ln(t)) \quad (13)$$

3 Estimate of maintenance efficiency

The object now is to study some estimators of maintenance efficiency since initial intensity is known. In that case the failure intensity is supposed to depend on a simple parameter $\rho \in \mathcal{J} : \lambda_t = \lambda_t(\rho)$. The true value of this parameter will be noted ρ_0 . Initially, the properties of maximum likelihood estimators (MLE) are studied and explicit estimators are introduced. We propose the following assumptions, appearing necessary for the treatment of this section.

- H_1 : $\mathcal{J}_0 = [\rho_1, \rho_2]$ is a known compact of \mathbb{R} , included inside \mathcal{J} and significant that $\rho_1 < \rho_0 < \rho_2$.
- H_2 : The functions $\lambda_{\min}, \lambda_{\max}$ are increasing, non null, and there exists a positive constant c , checking for all $\rho \in \mathcal{J}$ and $t \geq \gamma_1$: $\lambda_{\min}(t) \leq \lambda_t(\rho) \leq \lambda_{\max}(t)$ and $|\lambda'_t(\rho)| \leq c\lambda_{\min}(t)$, where $\lambda'_t(\rho)$ denotes the derivative in ρ of the restrictions on $] \gamma_1, +\infty[$ of the relations (1) and (2) respectively for the ARI_m and ARA_m reformulated models.
- H_3 : The MLE is required in a known interval : $\mathcal{J}_0 = [\rho_1, \rho_2]$ like that $-\infty < \rho_1 < \rho_0 < \rho_2 < 1$.

The MLE of maintenance efficiency, denoted $\hat{\rho}_t^{ML}$, is the value of ρ in \mathcal{J}_0 that maximizes likelihood. The convergence results for the MLE maintenance efficiency are given as follows:

For our reformulation of the ARI_m model and under assumptions H_1 and H_3 , the MLE of maintenance efficiency parameter, for only one observation of the failure process throughout interval $]\gamma_1, t]$, checks:

$$\sqrt{\frac{\Lambda(t)}{(1-\rho_0)^m}}(1-\rho_0)^m - (1-\hat{\rho}_t^{ML})^m \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (14)$$

In the same way as previously, and under the assumptions H_2 and H_3 , the MLE of maintenance efficiency parameter of the ARA_m reformulated model, for only one observation of the failure process in the interval $]\gamma_1, t]$, proves:

$$\sqrt{\frac{(t-\gamma_1)^{\beta_2}}{\eta_2(1-\rho_0)^{m(\beta_2-1)}}}(1-\rho_0)^{m(\beta_2-1)} - (1-\hat{\rho}_t^{ML})^{m(\beta_2-1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (15)$$

As it was seen according to preceding assumptions, we do not know how to prove that the MLE is convergent when the maximization of likelihood is made on $]-\infty, 1]$. So the MLE must be required in compact of $]-\infty, 1]$ containing the true value ρ_0 of maintenance efficiency. The explicit estimators (EE), which are not present in this problem type, can exist. These EE verify the same asymptotic properties as the MLE . They are based on the fact that the cumulative failure intensity can be appearing nearer asymptotically by the asymptotic intensity with an error proportional to a logarithm. The EE are also based on the fact that the cumulative failure intensity verifies:

$$\text{for all } t > \gamma_1, \Lambda_t = (1-\rho_0)^m \Lambda_\infty(t) + o(\Lambda(t)) \quad (16)$$

For the ARI_m reformulated model and under the assumption H_1 , for only one observation of the failure process over $]\gamma_1, t]$, the EE of maintenance efficiency parameter is given by:

$$\hat{\rho}_t^E = 1 - \left[\frac{\mathcal{N}_t}{\Lambda(t)} \right]^{1/m} \quad (17)$$

Similarly with the ARI_m reformulated model, an EE can be defined for the ARA_m reformulated model using the initial intensity in bath-tub form. Thus, we define near last reformulation, under the H_2 assumption, for only one observation of the failure process over $]\gamma_1, t]$, the EE of maintenance efficiency parameter. This estimator is expressed by:

$$\hat{\rho}_t^E = 1 - \left[\frac{\eta_2 \mathcal{N}_t}{t^{\beta_2}} \right]^{1/[m(\beta_2-1)]} \quad (18)$$

These two EE verifies the same convergence property as that of MLE (given respectively by the relations(16) and (17)). It is noticed that in particular case of the ARI_1 model, there exist an EE built from the true value of failure intensity. This estimator has the same

properties as the other estimators. And considering that it is not taken into account in term of its asymptotic development, we should hope that this estimator converges more quickly than the other estimators. We find in Doyen [9] that this estimator is expressed under the assumptions H_1 and H_3 , for only one observation of the failure process on $]\gamma_1, t]$, by the following relation:

$$\hat{\rho}_t^E = \frac{\Lambda(t) - \mathcal{N}_t}{\int_{\gamma_1}^t \lambda(\mathcal{T}_{\mathcal{N}_s}) ds} \quad (19)$$

proving the same convergence property as that of the MLE with $m = 1$.

In consideration to asymptotic normality of the estimators introduced in front, we maintain to define the Asymptotic Confidence Intervals (ACI). It is clear that for a same model, MLE and EE verify the same properties, then they describe the identical ACI . Thus, we can assimilate to the model ARI_m reformulated model with finished memory, simultaneously for two estimators under the assumptions H_1 , H_2 and H_3 , an ACI for $(1 - \rho_0)^m$ at level δ , given by:

$$ACI(\rho) = (1 - \hat{\rho})^m + \frac{u_\delta^2 \pm \sqrt{u_\delta^2 [4\Lambda(t)(1 - \hat{\rho})^m + u_\delta^2]}}{2\Lambda(t)}$$

where u_δ indicate it $1 - \frac{\delta}{2}$ quantile of the reduced-centered normal law, $\hat{\rho}$ indicate the MLE or EE and $\Lambda_{] \gamma_1, +\infty[}(t) = \frac{1}{\eta_0} t + \left(\frac{t - \gamma_1}{\eta_2}\right)^{\beta_2}$.

In a similar way, we can define an ACI for the ARA_m reformulated model. Under the assumptions H_1 and H_3 , the ACI for $(1 - \rho_0)^m$ of level δ is defined as follows:

$$ACI(\rho) = (1 - \hat{\rho})^{m(\beta_2 - 1)} + \frac{\eta_2 \left(u_\delta^2 \pm \sqrt{u_\delta^2 \left[\frac{4}{\eta_2} (t - \lambda_1)^{\beta_2} (1 - \hat{\rho})^{m(\beta_2 - 1)} + u_\delta^2 \right]} \right)}{2(t - \lambda_1)^{\beta_2}}$$

3.1 Simulation phase

Using simulations groping of the ARI_m and ARA_m models one next to one and for a given ACI , we estimate the coverage Rate (CR). This rate is expressed as the simulations proportion for which the true value of the parameter is in the confidence interval. Obviously, the CR converges to $1 - \delta$ when the number of observed failures n increases, where δ represents the ACI threshold. Practically, the CR is a function only of the estimator quality used to build the ACI . This is why we choose the CR as a comparison criterion for various estimators, rather than the mean squared error as it was the case in Doyen-Gaudoin [10]. For this fact, we have estimated over 10000 simulations with an initial intensity defined by the equation below, the CR of the ACI at level 95 for $m = 1, 2$ or 3 , $\beta_1 = 0.25$ or 0.75 , $\beta_2 = 1.5$ or 3 ,

$\rho = -1, -0.7, -0.5, -0.2, 0, 0.2, 0.5, 0.7$ or 0.9 and $n = 5, 10, 20, 40, 60, 80$ or 100 .

$$\lambda(t) = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t-\gamma_1}{\eta_2} \right)^{\beta_2-1} & \text{if } t > \gamma_1 \end{cases} \quad (20)$$

As it was often the study of empirical case of the Weibull model, the location and scale parameters γ and η seem not to have influence on the CR value. Thus, the graphs are the same ones for various parameters values $\gamma_0, \gamma_1, \eta_0, \eta_1$ and η_2 . The figures from 2 to 5 represent the estimators CR according to number n of simulated failures, for different values of β_1 and β_2 , the efficiency parameter ρ is equal to 0.5 .

The following notations are used in the various figures:

- | | | |
|----------|-----------|-----------|
| — MLE | ○ ARA_1 | * ARI_1 |
| ... EE | □ ARA_2 | + ARI_2 |
| | × ARA_3 | ▽ ARI_3 |

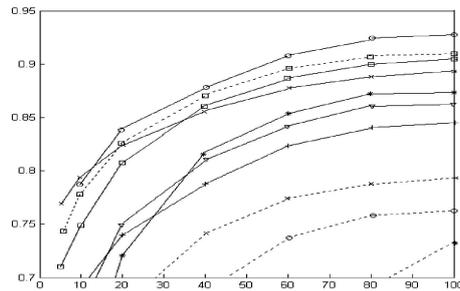


Figure 2: $CR(n); \beta_1 = 0.25; \beta_2 = 1.5$

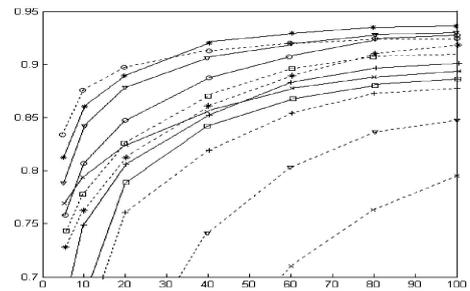


Figure 3: $CR(n); \beta_1 = 0.25; \beta_2 = 3$

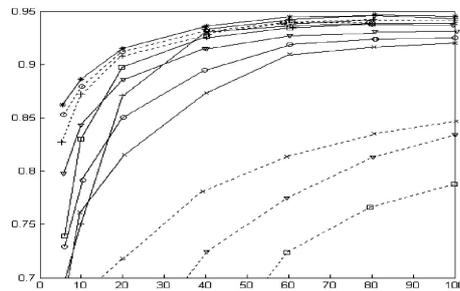


Figure 4: $CR(n); \beta_1 = 0.75; \beta_2 = 1.5$

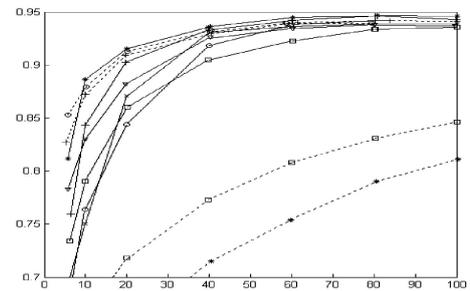


Figure 5: $CR(n); \beta_1 = 0.75; \beta_2 = 3$

It is noticed that for the MLE , in the case of decreasing phase (which is always convex in our model) of initial intensity is more convex (the value of the shape parameter β_1 closer

to zero, i.e. the system improves more quickly), the CR converges slowly for diverse models (figures 2 and 3). This incident can be explained by the fact that the number of failures occurring in the improvement period is proportional to the convexity degree of the initial intensity. Whereas if the increasing phase of the initial intensity is concave, the CR of ARI_m reformulated models converges less quickly than that one of ARA_m reformulated models, and vice versa if the increasing phase is convex. These results are respectively illustrated by the two figures 4 ($\beta_2 = 1.5$) and 5 ($\beta_2 = 3$). The idea is that in the convexity case ($\beta_2 > 2$), while making increase n and any thing remaining equal otherwise, the ARI_m reformulated models correspond to systems which are degraded more quickly than for those ARA_m . Thus, overall for the same number of observed failures, it is better to function the system according to the ARA_m reformulated model longer. Then, we have more information for the model in order to estimate its maintenance efficiency.

We see that for some EE the convergence speed, in the majority of cases, is very slow. Several of them having a CR values can be not read in the graphs, i.e. they are very small. The property of the MLE seems to be reversed for the EE . Moreover, since the EE are built with rapprochement principle of the failure intensity to the cumulative asymptotic intensity, then the EE tends to converge for enough large n values. Like the ARI_m reformulated model corresponds to system which is degraded more quickly than that ARA_m one, its cumulative and asymptotic intensities are near relatives, consequently, the EE converge more quickly.

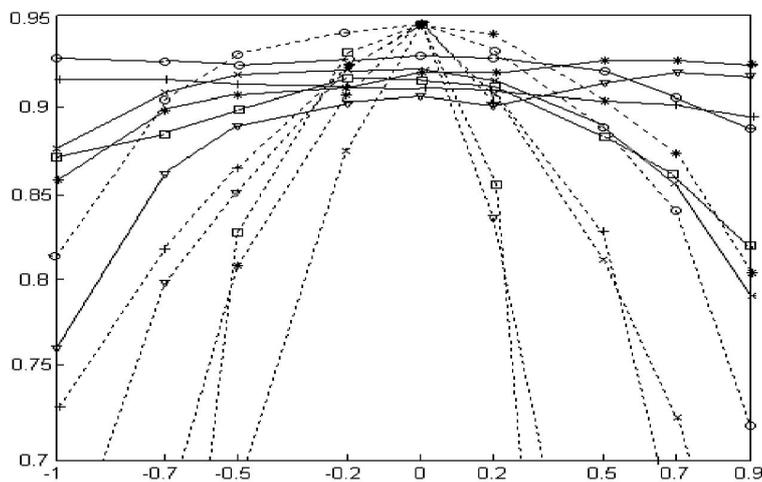


Figure 6: $CR(\rho)$; $\beta_1 = 0.75$; $\beta_2 = 3$; $n = 60$

The figure 6 represents the CR evolution according to the value of maintenance efficiency parameter ρ . So the CR of EE depends closely of the ρ value. The EE provides the most correct ACI for maintenance efficiency close to the ABAO case (ρ near to 0), even for a low values of the number of observed failures. On the other hand when maintenance efficiency

is too different to the ABAO assumption, the CR converges less quickly especially in the EE case. This result is a consequence due to the EE , which is founded on an equivalence property between cumulative failure and asymptotic intensities. This equivalence relation is made with a near remainder which is asymptotically equal to:

$$r(t) = \frac{(\beta_2 - 1)}{(1 - \rho)^m} \frac{1 - (1 + m\rho)(1 - \rho)^m}{\rho} \ln(t - \gamma_1) \quad (21)$$

At a certain instant $t > \gamma_1$ and for $\rho = 0$, this quantity is null by hypothesis. In fact in this case $\Lambda_t = \Lambda_\infty(t)$. whereas, when ρ tends to 1 by lower values, the above difference diverges. finally, since the maintenance efficiency is degraded and becomes more and more harmful, the cumulative asymptotic intensity increase, and the difference $r(t)$ tends to a constant limit equal to $m(\beta_2 - 1)$.

The MLE are characterized by CR , which are less sensitive to the value of ρ , but it is always under the assumption of ABAO maintenance efficiency that the estimators are most correct. This CR behavior whether through of the MLE or EE , can be owed to the operation of the system in the improvement and service life periods, and is maintained by ABAO maintenance actions. It appears clearly, on the one hand for low numbers of failures, and on the other hand for the models with high enough memories. Thus, for a great number of failures the ACI are the good approximations for the practical value of maintenance efficiency.

4 Conclusion

In this paper we reformulated two classes of imperfect maintenance models using failure intensity in bath-tub shape. We gave new results on our reformulations of arithmetic reduction of age or intensity with memory m . In fact, their failures process is characterized by equivalence between cumulative failure intensity and cumulative asymptotic intensity.

For the ARI and ARA reformulated models with finished memory, we proposed the explicit estimators of maintenance efficiency parameter. Then, we presented theoretical statistical results for the estimate of maintenance efficiency. The convergence properties relative to maximum likelihood and explicit estimators were derived. Thus, we could deduce that the asymptotic confidence intervals are issued from those estimators.

An eventual prospect for this work would be the theoretical study of simultaneous estimate of the initial intensity parameters with bath-tub form. The theoretical study of simultaneous estimate of all parameters does not seem indeed realistic.

References

- [1] Ascher H. (1990). A survey of tests for exponentiality. *Communications in Statistics-Theory and Methods*, 19, 1811-1825.
- [2] Ascher H. and Feingold H. (1984). Repairable Systems Reliability. *New York and Basel, Marcel Dekker*.
- [3] Baxter L., Kijima M., and Tortorella M. (1996), A point process model for the reliability of a maintained system subject to general repair, *Communications in Statistics*, 12, 37-65.
- [4] Block H., Borges W.S. et Savits T.H. (1985), Age-dependent minimal repair, *Journal of Applied Probability*, 22, 370-385.
- [5] Brown M. et Proschan F. (1983), Imperfect Repair, *Journal of Applied Probability*, 20, 851-859.
- [6] Cohen A. and Sacrowitz H.B. (1993). Evaluating Test for Increasing Intensity of a Poisson Process. *Technometrics*, 35, 446-458.
- [7] Dorado C., Hollander M. and Sethuraman J. (1997). Nonparametric estimation for a general repair model. *The Annals of Statistics* 25, 3, 1140-1160.
- [8] Doyen L. (2004). Estimation of maintenance efficiency in imperfect repair models. *4th International conference on Mathematical Methods in Reliability* (MMR 04), Santa Fe, USA, Juin 2004.
- [9] Doyen L. (2005). Repair efficiency estimation in the ARI_1 imperfect repair model. *Modern Statistical and Mathematical Methods in Reliability*, S. Keller-Mc Nulty, N. Limnios, A. Wilson and Y. Armijo eds., *World Scientific, Singapore*, 153-168.
- [10] Doyen L. and Gaudoin O. (2004), Classes of imperfect repair models based on reduction of failure intensity or virtual age, *Reliability Engineering and System Safety*, 84, 45-56.
- [11] Embrechts P., Klüppelberg C. and Mikosch T. (1997), Modeling extremal events, *Springer*.
- [12] Friedman L. and Gertsbakh I. (1980). Maximum Likelihood Estimation in a Minimum-type Model with Exponential and Weibull Failure Modes. *J.A.S.A*, 75, 460-465.
- [13] Jiang R., Ji P. and Xiao X. (2003). Aging property of unimodel failure rate models. *Reliability Engineering and System Safety*, 79, 1, 113-116.

- [14] Kijima M. (1989), Some results for repairable systems with general repair, *Journal of Applied Probability*, 26, 89-102.
- [15] Krit M., Rebaï A. and Benbacha H. (2007), General parametric reliability model for repairable system, To appear in *Reliability Engineering and System Safety*, 2008.
- [16] Last G. et Szekli R. (1998), Asymptotic and monotonicity properties of some repairable systems, *Advances in Applied Probability*, 30, 1089-1110.
- [17] Pham H. et Wang H. (1996), Imperfect maintenance, *European Journal of Operational Research*, 94, 452-438.
- [18] Shin I., Lim T.J. and Lie C.H. (1996), Estimating parameters of intensity function and maintenance effect for repairable unit, *Reliability Engineering and System Safety*, 57, 1-10.
- [19] Yun W.Y. and Choung S.J. (1999), Estimating maintenance effect and parameters of intensity function for improvement maintenance model, *Proc. 5th ISSAT Int. Conf. on Reliability and Quality in Design*, Las Vegas, 164-166.

Brown-Proschan imperfect repair model with bathtub failure intensity

Makrem KRIT * Abdelwaheb REBAÏ †

Abstract

The aim of this paper is to study the estimation of maintenance efficiency in Brown-Proschan model. This model has been proposed by Brown and Proschan (1983), the failure process is simply a Non Homogeneous Poisson Process (NHPP). Our model is defined by BP reformulation one using bathtub failure intensity. This form of intensity is presented like superposition of two NHPP and Homogeneous Poisson one (HPP). Moreover, the particularity of this model allows to take account of system state improvement in time course. The characteristics of failure process and its influence on maintenance process are studied while basing on Monte-Carlo simulation. Finally, the main features of our model are derived: The likelihood function, thus parameter estimation and evaluation of maintenance efficiency are possible.

Key words: repairable system, reliability, bathtub failure intensity, virtual age, imperfect maintenance, estimation, likelihood.

1 Introduction

The totality of the significant industrial systems is subjected to the actions corrective and preventive maintenance which are supposed to prolong their functional life. The efficiency evaluation of these maintenance actions is of a great practical interest, but it was seldom studied. In the literature, several models of maintenance effect were proposed. That is to say for example, Pham-Wang (1996) and Baxter-Kijima-Tortorella (1996). The authors tried to classify various models of maintenance. The majority of these models consider only the corrective maintenance (CM) effect, known under the name of repair models. These models are useful to model the real systems which are supported by a constant repair. Several repair models, including those of Brown-Proschan, the Block-Borges-Savits model (1985), the Kijima model (1989), the most general models of Dorado-Hollander-Sethuraman (1997) and Last-Szekli model (1998), were all useful

*Institut des Etudes Technologiques de Gafsa

†Ecole Supérieure de Commerce de Sfax

in this respect. Several theoretical properties, as well as the parameters estimators of fundamental failure intensity and their asymptotic intervals confidence studied by these authors, without evaluating the maintenance efficiency. The same claims of these models can be also used for the only preventive maintenance (PM).

The idea of the Brown-Proschan model (1983) is that, the efficiency of the k^{th} maintenance action is evaluated by a random variable E_k , independently and identically distributed according to the Bernoulli law with parameter p , such as:

$$E_k = \begin{cases} 1 & \text{if the maintenance is perfect} \\ 0 & \text{if the maintenance is minimal} \end{cases}$$

knowing that maintenance is always minimal over all the improvement period and that of service life (i.e. for $\mathcal{T}_k \leq \gamma_1$). We can show that at the moment t higher than γ_1 , the duration passed since the last perfect maintenance (moreover, before the moment γ_1 , all maintenances are supposed to be minimal) can be expressed in the form: $t - \mathcal{T}_{\mathcal{N}_t} + \sum_{k=\mathcal{N}_{\gamma_1}+1}^{\mathcal{N}_t} \left[\prod_{z=k}^{\mathcal{N}_t} (1 - E_z) \right] \mathcal{X}_k$, where $\mathcal{N}_{\gamma_1} = i + j$ represent the failures number (of maintenance action) will take place during, respectively the improvement and service life periods : i.e. before the instant γ_1 . Under these conditions, the failure intensity is written:

$$\lambda_t(\mathcal{N}, E) = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{1}{\eta_0} + \frac{\beta_2}{\eta_2} \left(\frac{t - \mathcal{T}_{\mathcal{N}_t} + \sum_{k=\mathcal{N}_{\gamma_1}+1}^{\mathcal{N}_t} \left[\prod_{z=k}^{\mathcal{N}_t} (1 - E_z) \right] \mathcal{X}_k}{\eta_2} \right)^{\beta_2-1} & \text{if } t > \gamma_1 \end{cases} \quad (1)$$

being given that the virtual age just after the k^{th} maintenance, noted a_k , is equal to $\sum_{z=\mathcal{N}_{\gamma_1}+1}^k \left[\prod_{h=z}^k (1 - E_h) \right] \mathcal{X}_h$ for all $k \geq \mathcal{N}_{\gamma_1} + 1$, where the variable \mathcal{X}_h indicate the h^{th} duration of the between-failures.

The figure1 translated the trajectory of this intensity for an unspecified value of p between 0 and 1. In this figure, the instants of perfect and minimal maintenances are represented on the x-axis respectively by circles and squares.

Concerning the evaluation of the maintenance efficiency, we return to the same properties presented in Brown and Proschan (1983). The reformulation of the Block-Borges-Savits model is also a generalization of the preceding form of the Brown-Proschan model. Indeed, the probability that the CM is perfect depends to the instant to which is carried out. In this case, the failure intensity is equivalent to that of the Brown-Proschan model, with that the law of E_k is related to parameters $p(\mathcal{T}_k)$.

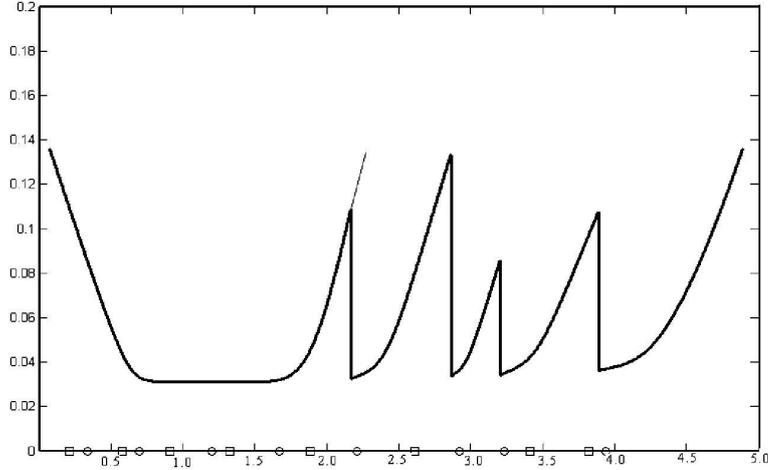


Figure 1: Reformulation of the Brown-Proschan model intensity

2 Characteristics of the failures process

The model Brown-Proschan (BP) is a particular case of the Kijima model (Kijima (1989)), making the two part of the whole models of repair. A generalization of such a model arises in Last-Szekli (1998) shows that the failures process, under certain conditions, tends to be stabilized. For our BP reformulated model, and in the degradation phase (obviously, the failure intensity is monotonous), the convergence property in law of the virtual age after maintenance (or effective age) and the waiting durations between two failures, is checked. This fact, the virtual age just after the k^{th} maintenance, noted a_k , have a distribution function of the form:

$$\mathcal{F}_{a_k}(t) = \begin{cases} 1 - \exp\left\{-\frac{1}{\eta_0}t\right\} \times \exp\left\{-\left(\frac{t}{\eta_1}\right)^{\beta_1} + \left(\frac{\gamma_0}{\eta_1}\right)^{\beta_1-1}t\right\} & \text{if } 0 < t < \gamma_0 \\ 1 - \exp\left\{-\frac{1}{\eta_0}t\right\} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ 1 - (1-p) \exp\{-\Lambda(t)\} \left[\sum_{z=\mathcal{N}_{\gamma_1+1}}^{k-1} (1-p)^z \frac{\Lambda^z(t)}{z!} \right] & \text{if } t > \gamma_1 \end{cases} \quad (2)$$

During the degradation period, and for a value of p strictly higher than zero, the random variables continuation of the ages $\{a_k\}_{k \geq \mathcal{N}_{\gamma_1+1}}$ converge in law towards a random variable a . Brown and Proschan (1983) proved that this variable follows a law having the failure rate $p\lambda(t)$, and $\mathcal{F}_a(t) = 1 - (1-p) \exp\{-\Lambda(t)\}$ as function of distribution. It is the same, as constantly, the virtual age is equal to the time passed since the last perfect maintenance. In Last-Szekli (1998), the authors showed the convergence of the continuation of the virtual ages expectation, and of the expectation of the between-failures durations. Within the framework of our BP reformulated model, and thanks to the convergence property of Brown-Proschan, we can obviously calculate these expectations and prove their tendency towards a finished and continuous limit.

The reformulation of the BP model using the failure intensity with bathtub form, characterize the failures process by the between-failures durations, for $k \geq \mathcal{N}_{\gamma_1} + 1$, having the survival function as:

$$\mathcal{S}_{\mathcal{X}_k}(x) = (1-p) \int_{\gamma_1}^{+\infty} \lambda(u) \exp\{-\Lambda(t+u)\} \left[(1-p)^{k-2} \frac{\Lambda^{k-2}(u)}{(k-2)!} + p \sum_{z=\mathcal{N}_{\gamma_1}}^{k-3} (1-p)^z \frac{\Lambda^z(u)}{z!} \right] du + p \exp\{-\Lambda(x)\} \quad (3)$$

And if moreover p is strictly positive, and that $\exp\{-\Lambda(x)\} = o(1)$, then the survival function of the between-failures durations, for the values of $k \geq \mathcal{N}_{\gamma_1} + 2$, converge in law towards the random variable \mathcal{X} with survival function:

$$\mathcal{S}_{\mathcal{X}}(x) = p \int_{\gamma_1}^{+\infty} \lambda(x+u) \exp\{-\Lambda(x+u) + (1-p)\Lambda(u)\} du \quad (4)$$

Thus, we deduce from these results, that the between-failures durations converge in law towards the random variable \mathcal{X} of which the survival function as:

$$\mathcal{S}_{\mathcal{X}}(x) = p \exp\{-\Lambda(x)\} + p(1-p) \int_{\gamma_1}^{+\infty} \lambda(u) \exp\{-\Lambda(x+u) + (1-p)\Lambda(u)\} du$$

Consequently, by using an integration by parts, we can write:

$$\begin{aligned} \mathcal{S}_{\mathcal{X}}(x) &= \lim_{u \rightarrow +\infty} \exp\{-\Lambda(x+u) + (1-p)\Lambda(u)\} \\ &\quad + p \int_{\gamma_1}^{+\infty} \lambda(x+u) \exp\{-\Lambda(x+u) + (1-p)\Lambda(u)\} du \end{aligned}$$

And considering that Λ is an increasing function, then:

$$\exp\{-\Lambda(x+u) + (1-p)\Lambda(u)\} \leq \exp\{-p\Lambda(u)\} = o(1)$$

Under these conditions, for our reformulation of the BP model by an intensity with bathtub form, if it exists an $\varepsilon > 0$ so that : $x^{1+\varepsilon} \exp\{-\Lambda(x)\} = o(1)$, then the expectation of the average waiting duration of the $(k+1)^{th}$ failure, with $k \geq \mathcal{N}_{\gamma_1} + 1$, is expressed by the following relation:

$$\mathbb{E}[x_k] = \int_{\gamma_1}^{+\infty} \exp\{-\Lambda(t)\} \left[(1-p)^{k-1} \frac{\Lambda^{k-1}(t)}{(k-1)!} + p \sum_{z=\mathcal{N}_{\gamma_1}}^{k-2} (1-p)^z \frac{\Lambda^z(t)}{z!} \right] dt \quad (5)$$

It appears that the most significant result of this section is, in the absence of any ambiguity, the relation (3) which offers the marginal laws of the consecutive between-failures durations. The figure 2 represents the simulated random rates of these between-failures durations of which the initial failure intensity in bathtub form (defined by the parameters $\eta_0 = 26$, $\eta_1 = \eta_2 = 200$, $\beta_1 = 0.5$, $\beta_2 = 3$, $\gamma_0 = 0.7$, $\gamma_1 = 1.6$) and a maintenance efficiency equalizes to 0.3.

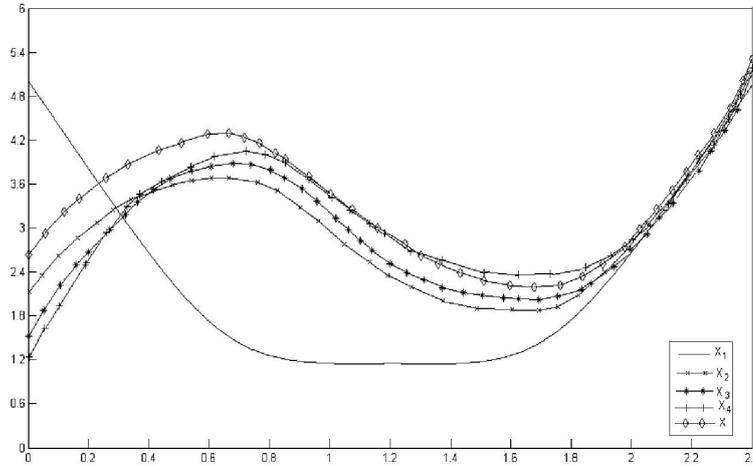


Figure 2: Random rate of the consecutive between-failures durations of the BP reformulated model

We notice that after the maintenance action, the random rate of the next between-failures duration is represented by a concave trajectory during the improvement period, and convex during the degradation period. The growth in k of the first values of this random rate is due to the fact that the effect of this maintenance action on the system is unknown. In other word, just after a maintenance action, the system tends moreover to weaken because what we don't know if the maintenance be effective or not. After being maintained, if the system survives long enough, the random rate takes the values almost identical to the initial intensity (it's extremely probable that this last maintenance is perfect).

The chart of the asymptotic random rate of the between-failures durations (simulated with the same parameters values of the initial failure intensity of Fig.2) (Fig.3) illustrate the maintenance effect. It's obviously the form of the random rate associated to the variable \mathcal{X} , already defined in the preceding paragraph, for various values of the efficiency parameter p .

We see that during all the period before the beginning of the degradation phase, the form of the random rate keeps always the same pace whatever is the parameter value p . In fact the maintenace effect is, by hypothesis, As Bad As Old (ABAO), and the maintenance action is carried out just to take again the system operation). The maintenance effect, through the degradation period, is to increase the first values of the random rate compared to that of the initial intensity. This increase is even higher than the parameter p is weak. So, more p is weak and more the maintenance is extremely probable that it's ABAO.

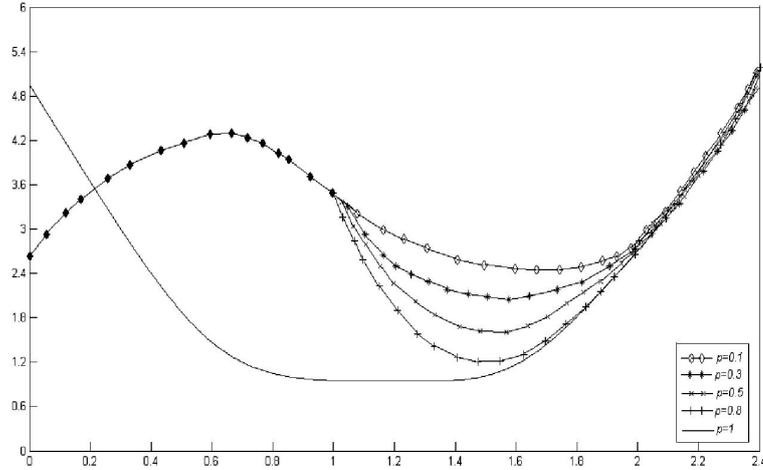


Figure 3: Asymptotic random rate of the between-failures durations according to various values of p

3 Estimate of the maintenance efficiency

Under the assumption that the maintenance effects are known (the maintenance is As Good As New (AGAN) or ABAO, and E_k are observed). Thus, the writing of the likelihood function is possible by using the equation (6). Therefore, we estimate the parameters of our model, such as the efficiency parameter p , and the parameters of the failure intensity λ . It's noticed that the estimator of p is logically the percentage of the perfect maintenance actions among all actions carried out. In the case of failure intensity of the Power-Law-Process type, Whitaker and Samagniego (1989) studied the identifiability problem of the parameter p for a waiting duration of the first failure according to the exponential law. And so the between-failures durations law is independent of the parameter p . Considering that the maintenance actions are useless when the system is neither in improvement state nor in degradation state, the models of assumptions AGAN and ABAO are the same ones.

$$\mathcal{L}(\theta; t_1, \dots, t_n) = \left[\prod_{i=1}^n \lambda_{t_i} \right] \exp \left\{ - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \lambda_s ds \right\} \quad (6)$$

If on the contrary, this first between-failure duration is supposed of non-exponential law, the identifiability problem does not arise and it's possible then to estimate in prior the parameter p . In the same way, we estimate the other parameters of the model, without having the observed values of the external variables E . Several alternatives of estimation, always for the simple failure intensity, namely the idea to use the Expectation Maximization (EM) algorithm, gotten by Lim (1998). In their article Lim, Lu and Park (1998) presented another method, based on the bayesian analysis, and which is trying to

give to p the prior law of the beta type. Another approach was proposed by Langseth and Lindqvist (2003), acted to calculate the model characterization without utilizing the external variables values E .

In practice, and without maintenance, the systems are considered either in improvement or in degradation states (in view of the service life phase is practically short). Then, the initial intensity cannot coincide with an exponential law and the identifiability problem doesn't presented. Indeed, this is logical insofar as the exponential law is in mental blank. Nevertheless, if the first between-failure duration is supposed of non-exponential law, the maintenance efficiency parameter is identifiable.

We interest thereafter, in the estimation of the parameters for our BP reformulated model by an intensity with bathtub form. The failures process depends on the external variables continuation of which we don't know their values. Subsequently, and even if the failures process is influenced by the external variables E , this can be in any event, also considered with a self-excited punctual process. The two following relations express the relationship between the failure intensities.

$$\lambda_t(\mathcal{N}, E) = \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr(\mathcal{N}_{t+dt} - \mathcal{N}_t = 1/\mathcal{N}_t, \mathcal{T}_{\mathcal{N}_t}, E_{\mathcal{N}_t})$$

and

$$\lambda_t(\mathcal{N}) = \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr(\mathcal{N}_{t+dt} - \mathcal{N}_t = 1/\mathcal{N}_t, \mathcal{T}_{\mathcal{N}_t})$$

We find in Andersen, Borgan and Gill (1993) the innovation theorem, allowing to note the failure intensity quite simply by $\lambda_t = \lambda_t(\mathcal{N})$. As the failure process is considered as well as a self-excited punctual process. We can apply within the parametric approach framework various procedures of estimate. This process is characterized by a clean failure intensity which is calculated in an iterative way according to the intensity values and cumulative failure intensity at the preceding maintenance instants. This function is expressed by the following equation.

$$\lambda_t = \begin{cases} \frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} (t^{\beta_1-1} - \gamma_0^{\beta_1-1}) & \text{if } 0 < t < \gamma_0 \\ \frac{1}{\eta_0} & \text{if } \gamma_0 \leq t \leq \gamma_1 \\ \frac{-d}{dt} \left[\ln \left(\sum_{k=\mathcal{N}_{\gamma_1}}^{\mathcal{N}_t} p^{1\{k>\gamma_1\}} (1-p)^{\mathcal{N}_t-k} \left[\prod_{j=k+1}^{\mathcal{N}_t} \frac{\lambda(\mathcal{T}_j - \mathcal{T}_k)}{\lambda_{\mathcal{T}_j}} \right] e^{\{-\Lambda(t-\mathcal{T}_k) - \Lambda_{\mathcal{T}_k}\}} \right) \right] & \text{if } t > \gamma_1 \end{cases}$$

Obviously, through the degradation phase, this property formalizing the model BP. What is translated by the third restriction of the foregoing equation . For this fact, we condition the calculation compared to the instant of the last perfect maintenance action. Consequently, we note $E_{z,k}$, for $k \leq z$ and $k \geq \mathcal{N}_{\gamma_1}$, the next event : the k^{th} maintenance is AGAN and the following ones, until the z^{th} maintenance, are ABAO:

$$E_{z,k} = \left\{ \{E_k = 1\}, \{E_j = 0\}_{k < j \leq z} \right\}.$$

$$\text{For all } t \geq 0, \quad \Pr(\mathcal{X}_{n+1} \geq x/\mathcal{T}_n) = \exp \left\{ - \int_{\mathcal{T}_n}^{\mathcal{T}_n+x} \lambda_u du \right\} \quad (7)$$

When the law of the failure instants is influenced by the maintenances process, \mathcal{M} , and by the external variables, E , the failure intensity remains insufficient to characterize perfectly the failures process. For this fact, and using the formula of the probability law of next between-failures time, data by the equation (7), and of the innovation theorem (1993) it's possible to write:

$$\Pr(\mathcal{X}_{n+1} \geq x/\mathcal{T}_n) = \exp \left\{ - \int_{\mathcal{T}_n}^{\mathcal{T}_n+x} \mathbb{E} \left[\lambda_u^{\mathcal{N}}(\mathcal{N}, \mathcal{M}, E) / \mathcal{N}_u = n, \mathcal{T}_n \right] du \right\} \quad (8)$$

We can deduce from this last relation the law of between-failures times, knowing the history of the failures process. In our reformulation case of the BP model, and through the degradation period, the failures process checks:

$$\Pr(\mathcal{X}_{n+1} \geq x/\mathcal{T}_n) = \sum_{k=\mathcal{N}_{\gamma_1}}^n p^{\mathbf{1}\{k > \gamma_1\}} (1-p)^{n-k} \left[\prod_{j=k+1}^n \frac{\lambda(\mathcal{T}_j - \mathcal{T}_k)}{\lambda_{\mathcal{T}_j}} \right] e^{\{-\Lambda(x+\mathcal{T}_n-\mathcal{T}_k)+\Lambda_{\mathcal{T}_n}-\Lambda_{\mathcal{T}_k}\}}$$

Thus, during all its life, the system is characterized by a failures process, of which a law of the between-failures times is given by:

$$\Pr(\mathcal{X}_{n+1} \geq x/\mathcal{T}_n) = \begin{cases} \exp \left\{ -\frac{1}{\eta_0} x - \left(\frac{\mathcal{T}_n+x}{\eta_1} \right)^{\beta_1} + \left(\frac{\mathcal{T}_n}{\eta_1} \right)^{\beta_1} + \left(\frac{\gamma_0}{\eta_1} \right)^{\beta_1-1} x \right\} & \text{if } n < \mathcal{N}_{\gamma_0} \\ \exp \left\{ -\frac{1}{\eta_0} x \right\} & \text{if } \mathcal{N}_{\gamma_0} < n < \mathcal{N}_{\gamma_1} \\ \sum_{k=\mathcal{N}_{\gamma_1}}^n p^{\mathbf{1}\{k > \gamma_1\}} (1-p)^{n-k} \left[\prod_{j=k+1}^n \frac{\lambda(\mathcal{T}_j - \mathcal{T}_k)}{\lambda_{\mathcal{T}_j}} \right] e^{\{-\Lambda(x+\mathcal{T}_n-\mathcal{T}_k)+\Lambda_{\mathcal{T}_n}-\Lambda_{\mathcal{T}_k}\}} & \text{if } n > \mathcal{N}_{\gamma_1} \end{cases}$$

We hold to distinguish between this property and that from the survival function. The Interpretations of the two properties are considerably distinct. The property corresponding to the survival function transmitted the marginal law of the consecutive between-failures durations. What allows us to well understand the evolution of the failures process. Whereas, the property of the preceding equation presents the conditional law of the consecutive between-failures durations. These two laws entirely characterize the failures process. Remain the problem of complexity which meets us in their studies.

We find in the figure 4 our reformulation with bathtub form of the self-excited failure intensity (in full feature), λ_t , and the failure intensity relating to the external process $\lambda_t(\mathcal{N}, E)$. The AGAN maintenance actions are indicated on the instants axis by circles, and the ABAO maintenance actions by squares. Subsequent to a maintenance action, the self-excited failure intensity proves to be the form of the initial intensity (in bathtub

form). When the duration passed since the preceding maintenance action is sufficiently long, the self-excited failure intensity is equal to the initial intensity at the instant $t - \mathcal{T}_{N_i}$. In consequence of a maintenance action, the pace of the self-excited failure intensity is dependent on the history of the failures process. Indeed, at what time the previous between-failures duration is sufficiently high, the length of the improvement period is less significant.

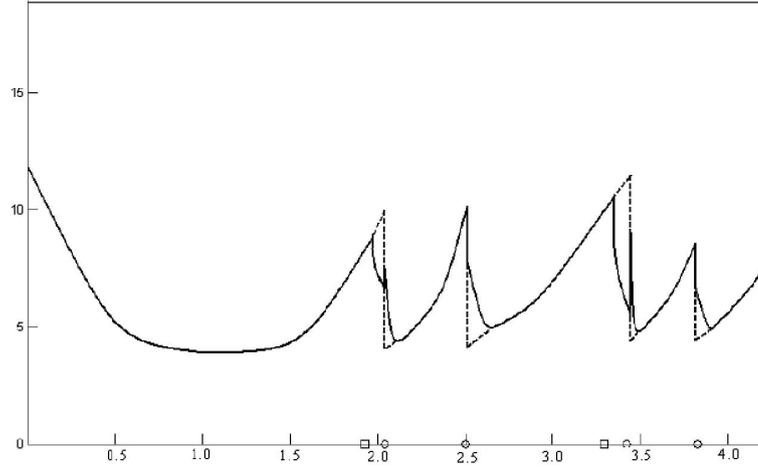


Figure 4: Reformulation in the bath-tub form of the self-excited and relative intensities of the BP model

By using the self-excited failure intensity defined, we can deduce the likelihood function associated to the observation of the maintenance instants. Under these conditions, our BP reformulated model allows to withdraw the likelihood function associated to only one observation of the failures process. By using the equation (6), this function is given by:

$$\begin{aligned} \mathcal{L}_{t_n}(\theta; t_1, \dots, t_n) &= \left[\prod_{k=1}^n \lambda_{\mathcal{T}_k} \right] \exp \left\{ - \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \lambda_s ds \right\} \\ &= \left[\prod_{k=1}^n \lambda_{\mathcal{T}_k} \right] \exp \{ - \Lambda_{t_n} \} \end{aligned}$$

This last function is equivalent to the likelihood function developed in the work of Doyen (2004). And that the restriction of this likelihood function over the two improvement and service life periods is equal to:

$$\begin{aligned} \mathcal{L}_{t_n\{n \leq \mathcal{N}_{\gamma_1}\}}(\theta; t_1, \dots, t_n) &= \left[\prod_{k=1}^i \left(\frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right) \right] \\ &\quad \times \left(\frac{1}{\eta_0} \right)^{n-i} \times \exp \left\{ - \left(\frac{\gamma_0}{\eta_1} \right)^{\beta_1} - \frac{1}{\eta_0} t_n \right\} \end{aligned}$$

Whereas through the degradation period, the restriction of this likelihood function is expressed as:

$$\mathcal{L}_{t_n\{n>\mathcal{N}_{\gamma_1}\}}(\theta; t_{\mathcal{N}_{\gamma_1}}, \dots, t_n) = \left[\prod_{k=\mathcal{N}_{\gamma_1}+1}^n \lambda_{\mathcal{T}_k} \right] \lambda_{t_n}^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} \exp \left\{ - \sum_{k=\mathcal{N}_{\gamma_1}+1}^n \int_{t_{k-1}}^{t_k} \lambda_s ds \right\}$$

Then it will be proved, in the following, that restriction likelihood function is written:

$$\begin{aligned} \mathcal{L}_{t_n\{n>\mathcal{N}_{\gamma_1}\}}(\theta; t_{\mathcal{N}_{\gamma_1}}, \dots, t_n) &= \sum_{k=\mathcal{N}_{\gamma_1}}^n p^{\mathbf{1}\{k>\gamma_1\}} \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n - \mathcal{T}_k) e^{-\Lambda(t_n - \mathcal{T}_k)} \\ &\quad (1-p)^{n-k} \left[\prod_{j=k+1}^n \lambda(\mathcal{T}_j - \mathcal{T}_k) \right] \left[\prod_{j=\mathcal{N}_{\gamma_1}+1}^k \lambda_{\mathcal{T}_j} \right] e^{-\Lambda_{\mathcal{T}_k}} \\ &= \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n) e^{-\Lambda(t_n)} (1-p)^n \left[\prod_{j=\mathcal{N}_{\gamma_1}+1}^n \lambda(t_n - \mathcal{T}_j) \right] + \sum_{k=\mathcal{N}_{\gamma_1}+1}^n p \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} \\ &\quad (t_n - \mathcal{T}_k) e^{-\Lambda(t_n - \mathcal{T}_k)} (1-p)^{n-j} \left[\prod_{j=k+1}^n \lambda(\mathcal{T}_j - \mathcal{T}_k) \right] \left[\prod_{j=\mathcal{N}_{\gamma_1}+1}^k \lambda_{\mathcal{T}_j} \right] e^{-\Lambda_{\mathcal{T}_k}} \end{aligned}$$

And owing to the fact that:

$$\mathcal{L}_{\mathcal{T}_k}(\theta; t_{\mathcal{N}_{\gamma_1}}, \dots, t_k) = \left[\prod_{j=1}^k \lambda_{\mathcal{T}_j} \right] \exp \{-\Lambda_{\mathcal{T}_k}\}$$

Therefore, while associating to the n^{th} observation, such that $n > \mathcal{N}_{\gamma_1}$, we obtain:

$$\begin{aligned} \mathcal{L}_{t_n\{n>\mathcal{N}_{\gamma_1}\}}(\theta; t_{\mathcal{N}_{\gamma_1}}, \dots, t_n) &= (1-p)^n \left[\prod_{j=\mathcal{N}_{\gamma_1}+1}^n \lambda(\mathcal{T}_j) \right] \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n) e^{-\Lambda(t_n)} \\ &+ p \left[\sum_{k=\mathcal{N}_{\gamma_1}+1}^n (1-p)^{n-k} \left(\prod_{j=k+1}^n \lambda(\mathcal{T}_j - \mathcal{T}_k) \right) \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n - \mathcal{T}_k) e^{-\Lambda(t_n - \mathcal{T}_k)} \mathcal{L}_{\mathcal{T}_k}(\theta) \right] \end{aligned}$$

Finally, the likelihood function of our reformulation of the BP model is overall definite by:

$$\mathcal{L}_{t_n}(\theta; t_1, \dots, t_n) = \mathcal{L}_{t_n\{n \leq \mathcal{N}_{\gamma_1}\}}(\theta; t_1, \dots, t_n) \times \mathcal{L}_{t_n\{n > \mathcal{N}_{\gamma_1}\}}(\theta; t_{\mathcal{N}_{\gamma_1}}, \dots, t_n) \quad (9)$$

It appears that in our study, the fact of removing the logarithm of the likelihood function doesn't simplify calculations. Thereafter, we attach to the direct calculation of the first partial derivative of likelihood in p . This calculation is resulted in the following function:

$$\begin{aligned} \frac{\partial}{\partial p} \mathcal{L}_{t_n}(\theta; t_1, \dots, t_n) &= \left[\prod_{k=1}^i \left(\frac{1}{\eta_0} + \frac{\beta_1}{\eta_1^{\beta_1}} \left(t_k^{\beta_1-1} - \gamma_0^{\beta_1-1} \right) \right) \right] \left(\frac{1}{\eta_0} \right)^{n-i} e^{-\left(\frac{\gamma_0}{\eta_1}\right)^{\beta_1} - \frac{1}{\eta_0} t_n} \times \left\{ -n(1-p)^{n-1} \right. \\ &\quad \left[\prod_{j=\mathcal{N}_{\gamma_1}+1}^n \lambda(\mathcal{T}_j) \right] \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n) e^{-\Lambda(t_n)} + \sum_{k=\mathcal{N}_{\gamma_1}+1}^n (1-p)^{n-k-1} \left(\prod_{j=k+1}^n \lambda(\mathcal{T}_j - \mathcal{T}_k) \right) \\ &\quad \lambda^{\mathbf{1}\{t_n=\mathcal{I}_{\mathcal{N}_{t_n}}\}} (t_n - \mathcal{T}_k) e^{-\Lambda(t_n - \mathcal{T}_k)} \left[(1 - (n - k + 1)p) \mathcal{L}_{\mathcal{T}_k}(\theta) + p(1-p) \frac{\partial}{\partial p} \mathcal{L}_{\mathcal{T}_k}(\theta) \right] \left. \right\} \end{aligned}$$

Given that the calculation of the partial derivative of likelihood in the parameters of the initial intensity is much more complex, we can then use their estimators without the maintenance process. Indeed, the estimate is carried out being given only the failures process, and the maintenance actions are supposed all minimal. Moreover, the two estimate procedures presented (the direct maximum likelihood and the EM algorithm) get in their globally the best estimators, especially that of the EM algorithm.

4 Conclusion

In this study, we gave new results on our new reformulation of the Brown-Proschan model. Doyen (2004) proved that this model corresponds to systems for which the maintenance efficiency makes it possible to contain or to stabilize degradation. That enabled us to introduce the innovation theorem which makes it possible to treat our general model of maintenance efficiency including the hidden external variables, in a way similar to a self-excited punctual process.

In the simulation stage, it's noticed that the service life period was not taken into account, in the direction where the system state is stabilized during this period. Moreover, in practice the service life period is, in general, short compared to the total life period of the reparable systems.

We can then, for this general model, to calculate a failure intensity known as clean and a clean likelihood function. Within the framework of the BP model, this failure intensity and this clean likelihood are complex and must be calculated recursively. Moreover, in spite of the complexity of the clean likelihood function, we showed that it was possible to calculate its partial derivative and thus to maximize it by numerical methods.

By knowing the clean likelihood function, other methods of estimate can be used. Like a method of Newton or via groping by calculating all the clean likelihood values on a grid.

References

- [1] Andersen P.K., Borgan O. and Gill R.D. (1993), Statistical models based on counting processes, *Springer Series in Statistics, Springer-Verlag*.
- [2] Baxter L., Kijima M., and Tortorella M. (1996), A point process model for the reliability of a maintained system subject to general repair, *Communications in Statistics*, 12, 37-65.

- [3] Block H., Borges W.S. and Savits T.H. (1985), Age-dependent minimal repair, *Journal of Applied Probability*, 22, 370-385.
- [4] Brown M. and Proschan F. (1983), Imperfect Repair, *Journal of Applied Probability*, 20, 851-859.
- [5] Kijima M. (1989), Some results for repairable systems with general repair, *Journal of Applied Probability*, 26, 89-102.
- [6] Dorado C., Hollander M. and Sethuraman J. (1997), Nonparametric estimation for a general repair model. *The Annals of Statistics* 25, 3, 1140-1160.
- [7] Doyen L. (2004), Estimation of Maintenance Efficiency in Imperfect Repair Models, *Mathematical Methods in Reliability (MMR 2004)*, Santa Fe, USA, June 2004.
- [8] Langseth H. and Lindqvist B.H. (2003), A maintenance model for components exposed to several failure mechanisms and imperfect repair, *Quality, Reliability and Engineering Statistics*. World Scientific Publishing.
- [9] Last G. and Szekli R. (1998), Asymptotic and monotonicity properties of some repairable systems, *Advances in Applied Probability*, 30, 1089-1110.
- [10] Lim T.J. (1998), Estimating system reliability with fully masked data under Brown-Proschan imperfect repair model, *Reliability Engineering and System Safety* 59, 277-289.
- [11] Lim T.J., Lu K.L. and Park D.H. (1998), Bayesian imperfect repair model, *Communications in Statistics-Theory and Methods*, 27(4), 965-984.
- [12] Pham H. and Wang H. (1996), Imperfect maintenance, *European Journal of Operational Research*, 94, 452-438.
- [13] Whitaker L.R. and Samaniego F.J. (1989), Estimating the reliability of systems subject to imperfect repair, *Journal of the American Statistical Association*, 84, 301-309.

The Itô transform for a general class of pseudo-differential operators

Rémi Léandre
Institut de Mathématiques. Université de Bourgogne.
21000. Dijon. FRANCE.

January 15, 2010

Abstract

We give an Itô formula for a general class of pseudo-differential operators.

1 Introduction

Let us recall what is the Itô formula for a purely discontinuous martingale $t \rightarrow M_t$ with values in \mathbb{R} [1]. Let f be a C^2 function on \mathbb{R} . We have

$$f(M_t) = f(M_0) + \int_0^t f'(M_{s-}) \delta M_s + \sum_{s \leq t} f(M_s) - f(M_{s-}) - f'(M_{s-}) \Delta M_s \quad (1)$$

It is the generalization of the celebrated Itô formula for the Brownian motion $t \rightarrow B_t$ on \mathbb{R} [1]

$$f(B_t) = f(B_0) + \int_0^t f'(B_s) \delta B_s + 1/2 \int_0^t f''(B_s) ds \quad (2)$$

A lot of stochastic analysis tools for diffusions were translated by Léandre in semi-group theory in [6], [7], [8], [10], [11], [14], [15], [16], [18]. Some basical tools of stochastic analysis for the study of jump processes were translated by Léandre in semi-group theory in [11], [12], [19]. For review on that, we refer to the review of Léandre [9], [17].

Léandre has extended the Itô formula for the Brownian motion to the case of some classical partial differential equations in [19], [21], [22], [23]. In such a case, there is until now no convenient measure on a convenient path space associated to this partial differential equation. In [23], we have extended the Itô formula for jump process for an integro-differential generator when there is until now no stochastic process associated. Jump processes are generically generated by pseudo-differential operators which satisfy the maximum principle [5].

In this paper, we give an Itô formula for a general class of positive elliptic pseudo differential operators. For material on pseudo-differential operators, we refer on [2], [3], [4] and [5]. Since the considerations below on pseudo-differential operators are more and less classical, we won't enter in the technical details of the proof.

2 The two semi-groups

Let \hat{u} be the Fourier transform of a smooth function u on \mathbb{R}^d . Let $a(x, \xi)$ be a global symbol of order m on \mathbb{R}^d . It is a smooth function from $\mathbb{R}^d \times \mathbb{R}^d$ into \mathbb{C} such that for all k, k'

$$\sup_{x \in \mathbb{R}^d} |D_x^k D_\xi^{k'} a(x, \xi)| \leq C_{k, k'} |\xi|^{m-k'} \quad (3)$$

We say that a global symbol of order m is elliptic if for $|\xi| > M$

$$\inf_{x \in \mathbb{R}^d} |a(x, \xi)| \geq C_M |\xi|^m \quad (4)$$

We consider the proper pseudodifferential operator associated to the symbol a : the Fourier transform of $L_0 u$ is given by

$$\int_{\mathbb{R}^d} a(x, \xi) \hat{u}(\xi) d\xi \quad (5)$$

We consider its adjoint L_0^* on $L^2(dx)$ and we put $L = L_0^* L_0$.

All the considerations of [2] which were valid on a compact subset of \mathbb{R}^d are still true because (3) and (4) are valid globally. In particular, L is essentially selfadjoint on $L^2(dx)$ and generates a contraction semi-group P_t on $L^2(dx)$.

Let us consider a smooth function f from \mathbb{R}^d into \mathbb{R} with compact support and a smooth function v with compact support from $\mathbb{R}^d \times \mathbb{R}$ into \mathbb{C} . (x, y) denotes the generic element of $\mathbb{R}^d \times \mathbb{R}$. We consider the smooth function from \mathbb{R}^d into \mathbb{R} \hat{v}

$$\hat{v}(x) = v(x, f(x)) \quad (6)$$

We consider the function \bar{v} from $\mathbb{R}^d \times \mathbb{R}$ into \mathbb{C}

$$(x, y) \rightarrow v(x, y + f(x)) \quad (7)$$

We apply L to \bar{v} , y being frozen. We get a function $L\bar{v}$. We put

$$(\hat{L}v)(x, y) = (L\bar{v})(x, y - f(x)) \quad (8)$$

Definition 1 \hat{L} is called the Itô transform of L .

We remark that $(x, y) \rightarrow (x, y + f(x))$ is a diffeomorphism of $\mathbb{R}^d \times \mathbb{R}$ which keeps the measure $dx \otimes dy$ invariant. This shows:

Theorem 2 \hat{L} is positive symmetric on $L^2(dx \otimes dy)$. It admits therefore a self-adjoint extension still denoted \hat{L} . This self-adjoint extension generates a semigroup \hat{P}_t of contraction on $L^2(dx \otimes dy)$

We get

Theorem 3 (Itô formula) We have the relation for all smooth function v with compact support

$$P_t(\hat{v})(x) = (\hat{P}_t(v))(x, f(x)) \quad (9)$$

Remark: If we consider the generator $L = \sum X_i^2$ where the X_i are smooth vector fields, $\hat{L} = \sum \hat{X}_i^2$ where

$$\hat{X}_i = (X_i, \langle X_i, df \rangle) \quad (10)$$

which corresponds to the generator of [19], [21], [22]. Analogous remark holds for the considerations of [23].

3 Proof of the Itô formula

Lemma 4 If v is a smooth function on $\mathbb{R}^d \times \mathbb{R}$ whose all derivatives belong to L^2 , $\hat{P}_t v$ is still a smooth function whose all derivatives belong to L^2 .

Proof: Let

$$\bar{L} = \hat{L} + \left(-\frac{\partial^2}{\partial y^2}\right)^{m/2} \quad (11)$$

\bar{L} commute with \hat{L} . Therefore, for all k

$$(\bar{L}^k) \hat{P}_t = (\hat{P}_t)(\bar{L}^k) \quad (12)$$

If v satisfies the hypothesis, $\hat{P}_t v$ belongs to the domain of \bar{L}^k . But \bar{L} is the transform of

$$\tilde{L} = L + \left(-\frac{\partial^2}{\partial y^2}\right)^{m/2} \quad (13)$$

under the change of variable $(x, y) \rightarrow (x, y + f(x))$. Therefore $\hat{P}_t v$ belongs to the domain of \tilde{L}^k . The result arises by Garding inequality. \diamond

Let ϕ be a smooth function from \mathbb{R}^d into $[0, 1]$, equals to 0 if $|\xi| \geq 2$ and equals to 1 if $|\xi| \leq 1$. We consider the global symbol

$$a_\lambda(x, \xi) = \phi(\xi/\lambda) a(x, \xi) \quad (14)$$

and the operator $L_{0,\lambda}, L_{0,\lambda}^*$ associated to it. Classically

$$L_{0,\lambda} u(x) = \int_{\mathbb{R}^d} K_\lambda(x, y) u(y) dy \quad (15)$$

$$L_{0,\lambda}^* u(x) = \int_{\mathbb{R}^d} \bar{K}_\lambda(y, x) u(y) dy \quad (16)$$

Lemma 5 *If u is smooth whose all derivative belong to L^2 , then $(L_0 - L_{0,\lambda})u$ tends to zero as well as all his derivatives and in L^2 when $\lambda \rightarrow \infty$. The same holds for $(L_0^* - L_{0,\lambda}^*)u$.*

Proof: $(L_0 - L_{0,\lambda})u$ is given by the oscillatory integral

$$\int \int_{\mathbb{R} \times \mathbb{R}^d} \exp[2\pi i \langle x - y | \xi \rangle (1 - \phi(\xi/\lambda)) a(x, \xi) u(y) dy d\xi \quad (17)$$

The result holds by integrating by parts in y . Analog statement work for $(L_0^* - L_{0,\lambda}^*)u$. \diamond

Proof of the Itô formula: We put

$$L_\lambda = L_{0,\lambda}^* L_{0,\lambda} \quad (18)$$

L_λ is a continuous operator acting on bounded continuous function on \mathbb{R}^d endowed with its uniform norm. The same is true for its Itô transform \hat{L}_λ . Therefore L_λ generates a semi-group $P_{\lambda,t}$ on bounded continuous functions on \mathbb{R}^d . \hat{L}_λ generates a semi-group $\hat{P}_{\lambda,t}$ on bounded continuous functions on $\mathbb{R}^d \times \mathbb{R}$. Moreover if u and v are bounded continuous,

$$P_{\lambda,t} u = \sum 1/n! L_\lambda^n u \quad (19)$$

and

$$\hat{P}_{\lambda,t} v = \sum 1/n! \hat{L}_\lambda^n v \quad (20)$$

But

$$L_\lambda^n \hat{v}(x) = (\hat{L}_\lambda^n v)(x, f(x)) \quad (21)$$

Therefore

$$P_{\lambda,t} \hat{v}(x) = (\hat{P}_{\lambda,t} v)(x, f(x)) \quad (22)$$

But $(\hat{P}_{\lambda,t} - \hat{P}_t)(v)$ is solution of the parabolic equation

$$-d/dt v_t = \hat{L}_\lambda v_t + (\hat{L}_{\lambda,t} - \hat{L}) \hat{P}_t v \quad (23)$$

with initial condition 0. The result arises from the two previous lemma, by the method of variation of constants since $\hat{P}_{\lambda,t}$ is a semi-group of contraction on $L^2(dx \otimes dy)$. This shows that for $\lambda \rightarrow \infty$

$$\hat{P}_{\lambda,t} v \rightarrow \hat{P}_t v \quad (24)$$

in $L^2(dx \otimes dy)$. Similarly, in $L^2(dx)$

$$P_{\lambda,t} \hat{v} \rightarrow P_t \hat{v} \quad (25)$$

We remark that \hat{L}_λ commute with \bar{L} . Therefore

$$(\bar{L}^k)(\hat{P}_{\lambda,t} - \hat{P}_t)v = (\hat{P}_{\lambda,t} - \hat{P}_t)(\bar{L}^k v) \quad (26)$$

By a similar argument to the proof of lemma (4), we can show that the convergence in (24) and (25) works for the uniform topology and not in L^2 only. This shows the result. \diamond

References

- [1] C. Dellacherie, P.A. Meyer: Probabilités et potentiel (II). Théorie des martingales. Hermann. Paris (1980).
- [2] J. Dieudonné: Eléments d'analyse VII. Gauthier-Villars. Paris (1977).
- [3] P. Gilkey: Invariance theory, the heat equation and the Atiyah-Singer theorem. Sd edition. CRC Press, Boca Raton (1995).
- [4] L. Hörmander: The analysis of linear partial differential operators (III). Springer, Heidelberg (1984)
- [5] N. Jacob: Pseudo differential operators. Markov processes (II). Generators and their potential theory. Imperial College Press. London (2002).
- [6] R. Léandre: Malliavin Calculus of Bismut type without probability. In "Festschrift in honour of K. Sinha".A.M. Boutet de Monvel and al eds, Proc. Indian. Acad. Sci (Math. Sci), 116, 2006, 507-518. *arXiv:0707.2143v1[math.PR]*
- [7] R. Léandre: Varadhan estimates without probability: lower bounds. In "Mathematical methods in engineering" (Ankara), D. Baleanu and al eds. Springer, Heidelberg, 2007, 205-217.
- [8] R. Léandre: Positivity theorem in semi-group theory. Mathematische Zeitschrift 258, 2008, 893-914.
- [9] R. Léandre: Applications of the Malliavin Calculus of Bismut type without probability. In "Simulation, Modelling and Optimization" (Lisboa), A. M. Madureira C.D. 2006, pp. 559-564. WSEAS transactions on mathematics 5, 2006, 1205-1211.
- [10] R. Léandre: The division method in semi-group theory. In "Applied mathematics" (Dallas) K. Psarris edt. W.S.E.A.S. press, Athens, 2007, 7-11.
- [11] R. Léandre: Leading term of a hypoelliptic heat-kernel. WSEAS Transactions on mathematics 6, 2007, 755-763.
- [12] R. Léandre: Girsanov transformation for Poisson processes in semi-group theory. In "Num. Ana. Applied. Mathematics." (Corfu) T. Simos edt. A.I.P. Proceedings 936, 2007, 336-339.
- [13] R. Léandre: Malliavin Calculus of Bismut type for Poisson processes without probability. In "Fractional order systems". J. Sabatier and al eds. Jour. Eur. systemes Automatisés. 42, 2008, 715-733.
- [14] R. Léandre: Wentzel-Freidlin estimates in semi-group theory. in "Control, Automation, Robotics and Vision" (I.E.E.E.), (Hanoi), Yeng Chai Soh edt, C.D., I.E.E.E. 2008, 2233-2236.

- [15] R. Léandre: Varadhan estimates in semi-group theory: upper bound. "Applied computing conference" (Istanbul).M. Garcia-Planas and al eds. W.S.E.A.S. press, 2008, 77-80.
- [16] R. Léandre: Varadhan estimates without probability: upper bound. WSEAS transactions on mathematics 7, 2008, 244-253.
- [17] R. Léandre: Malliavin Calculus of Bismut type in semi-group theory. Far East Journal of Mathematical Sciences 30, 2008, 1-26.
- [18] R. Léandre: Large deviations estimates in semi-group theory. In "Num. Ana. Applied. Mathematics" (Kos), T. Simos edt. A.I.P. Proceedings 1048, 2008, 351-355.
- [19] R. Léandre: Itô-Stratonovitch formula for a four order operator on a torus. In "Non-Euclidean Geometry and its applications" (Debrecen), S. Nagy edt, Acta Physica Debrecina 42, 2008, 133-137.
- [20] R. Léandre: Regularity of a degenerated convolution semi-group without to use the Poisson process. To appear in "Nonlinear Science and Complexity" (Porto) M. Silva and al eds. Mittag Leffler Preprint. Fall 2007. S.P.D.E. 10.
- [21] R. Léandre: Itô-Stratonovitch formula for the Schroedinger equation associated to a big order operator on a torus. in "Fractional order differentiation" (Ankara), G. Zaslavsky and al eds, Physica Scripta T 136, 2009, 014028.
- [22] R. Léandre: Itô-Stratonovitch formula for the wave equation on a torus. To appear in "Computations of stochastic systems". M.A. El-Tawil edt.
- [23] R. Léandre: Itô formula for an integro differential operator without an associated stochastic process. To appear in "ISAAC 2009" (London), J. Wirth edt.
- [24] R. Léandre: Wentzel-Freidlin estimates for jump processes in semi-group theory: lower bound. Preprint.
- [25] R. Léandre: Wentzel-Freidlin estimates for jump processes: upper bound. To appear in "Worldcomp 10"(Las-Vegas) H. Arabnia edt
- [26] K Yosida: Functional analysis. Springer, Heidelberg, 1977.

Wentzel-Freidlin estimates in semi-group theory

Rémi Léandre

Institut de Mathématique de Bourgogne
Université de Bourgogne. 21000. Dijon. France
(e-mail: Remi.leandre@u-bourgogne.fr)

Abstract. We give the translation in semi-group theory of Wentzel-Freidlin estimates either for diffusion or jump processes. We give an application to Varadhan estimates for heat-kernel when we use a mixture of large deviation estimates and of the Malliavin Calculus of Bismut type translated by us in semi-group theory.

Keywords: Wentzel-Freidlin estimates, Varadhan estimates.

1 The case of a diffusion

In the first case we study the case of the small time behaviour of a diffusion. We translate in semi-group our proof of large deviation theory for diffusion which was valid for all the path space. Here this estimate is valid only for the semi-group. In particular, our proof is based upon the translation in semi-group theory of martingale exponential and of the Itô formula. By using a mixture between the Malliavin Calculus and large deviation theory, as it was pioneered by Bismut in his celebrated book "Large deviation and the Malliavin Calculus", we can translate in semi-group our proof of Varadhan estimate, upper bound, for a subelliptic heat-kernel which say when Wentzel-Freidlin estimate pass to heat-kernel.

2 The case of a Poisson process

In this case, we consider the behaviour of a jump process with more and more jumps which belong smaller and smaller. We show it is related to the theory of semi-classical expansion of Maslov and others people by looking the symbol of the operator. We translate in semi-group theory the proof of large deviation theorems of Wentzel-Freidlin which were valid for all the path space and now is valid only for the semi-group. We do a mixture between the Malliavin Calculus for jump process of Bismut type translated by ourself in semi-group theory and these Wentzel-Freidlin estimates in order to establish a logarithmic expansion of the involved heat-kernel.

References

1. Landre R., "Wentzel-Freidlin estimates in semi-group theory", In *Control, Automation, Robotics and Vision*, Yeng-Saih Choh and al eds, C.D.,I.E.E.E., 2233–2236, (2008).
2. Landre R., "Malliavin Calculus of Bismut type in semi-group theory", *F.J.M.S.* 30, 1–26 (2008).
3. Landre R., "Regularity of a degenerated convolution semi-group without to use the Poisson Process", To appear in "*Non linear science and complexity II*", M. Silva and al eds.
4. Landre R., "Malliavin Calculus of Bismut type without probability", In *Fractional order systems*", *J.E.S.A.*, 715–733, (2008).
5. Landre R., "Wentzel-Freidlin estimates for jump process in semi-group theory: lower bound", In *Int.Conf.Dif. Geom. Dynamical Systems*", V. Balan edt, *B.S.G. Proceedings 17*, 107–113, (2010).
6. Landre R., "Wentzel-Freidlin estimates for jump process in semi-group theory: upper bound", To appear *Conf.Scientific. Computing*, H. Arabnia edt, (2010).
7. Landre R., "Varadhan estimates for a degenerated convolution semi-group: upper bound". Preprint 596. Univ. Bourgogne (2010)

Models of Statistic Distributions of Nonparametric Goodness-of-fit Tests in Composite Hypotheses Testing in Case of Generalized Weibull Law

Boris Yu. Lemeshko, Stanislav B. Lemeshko and Kseniya A. Akushkina

Department of applied mathematics and computer science, Novosibirsk State Technical
 University
 Novosibirsk, Russia
 Email: skyer@mail.ru

Abstract: In this paper there are presented results (tables of percentage points and statistic distribution models) for the nonparametric goodness-of-fit tests in testing composite hypotheses using the maximum likelihood method for parameters estimation for Generalized Weibull Distribution law. Statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulation.

Keywords: Goodness-of-fit test, Composite hypotheses, Kolmogorov test, Cramer-Von Mises-Smirnov test, Anderson-Darling test, Generalized Weibull distribution.

1 The family of Generalized Weibull Distribution in Reliability

Let us consider the sample $X = (X_1, X_2, \dots, X_n)^T$, we say that X_i follow the generalized Weibull distribution (GWD). The density function of the law is defined by:

$$f(x; \theta_0, \theta_1, \theta_2) = \frac{\theta_0}{\theta_1} \theta_2^{\theta_0} x^{\theta_0-1} \left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}-1} e^{-\left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}}}, \quad (1)$$

where $x \geq 0$, $\theta_0, \theta_1, \theta_2 > 0$. The family (1) defines a set of different laws. Special cases of Generalized Weibull Distribution are: $\theta_1 = 1$ – the family of Weibull distribution; $\theta_0 = 1, \theta_1 = 1$ – the family of exponential distribution.

The distribution function is

$$F(x; \theta_0, \theta_1, \theta_2) = 1 - e^{-\left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}}}.$$

The hazard function of GWD can be monotone increase ($\theta_0 > 1, \theta_0 > \theta_1$ and $\theta_0 = 1, \theta_1 < 1$), monotone decrease ($0 < \theta_0 < 1, \theta_0 < \theta_1$ и $0 < \theta_0 < 1, \theta_0 = \theta_1$), \cap -shaped ($\theta_1 > \theta_0 > 1$), \cup -shaped ($0 < \theta_1 < \theta_0 < 1$) and we have:

$$\lambda(x) = \frac{\theta_0}{\theta_1} \theta_2^{\theta_1} x^{\theta_0-1} \left(1 + \left(\frac{x}{\theta_2} \right)^{\theta_0} \right)^{\frac{1}{\theta_1}-1}. \quad (2)$$

The Generalized Weibull distribution used in reliability and survival tasks along with lognormal distribution and inverse Gaussian distribution. As usual, in construction the models of laws for real observed random variables it is difficult to discriminate one laws from another and choose one of them. To certain degree these difficulties related with restricted facilities of application nonparametric goodness-of-fit tests with unknown statistic distribution for verification of composite hypotheses concerning GWD.

2 Nonparametric goodness-of-fit tests in verification single and composite hypotheses

One of the most popular statistic analysis problems in handling the results of experimental data is verification the agreement experimental distribution and theoretical one. There exist the verification of single hypothesis and composite hypothesis. Single hypothesis has the form $H_0: F(x) = F(x, \theta)$, $F(x, \theta)$ is probability distribution function, θ is known parameter value (θ is scalar parameter or vector parameter). In the case of single hypothesis marginal statistic distribution of nonparametric Kolmogorov, ω^2 Cramer-Von Mises-Smirnov, Ω^2 Anderson-Darling goodness-of-fit tests do not depend on view of observed distribution law and parameters values. These goodness-of-fit tests are "free from the distribution"

Composite hypotheses has the form $H_0: F(x) \in \{F(x, \theta), \theta \in \Theta\}$. In this case the estimate of distribution parameter $\hat{\theta}$ is calculated by the same sample, the nonparametric Kolmogorov, ω^2 Cramer-Von Mises-Smirnov, Ω^2 Anderson-Darling goodness-of-fit tests lose the property called "free from the distribution".

In Kolmogorov goodness-of-fit test the value

$$D_n = \sup_{|x| < \infty} |F_n(x) - F(x, \theta)|,$$

where $F_n(x)$ is the empirical distribution function, n is the sample size, is used in Kolmogorov test characterized a distance between the empirical and theoretical laws. In testing of hypotheses used a statistic with Bolshev correction [1] in the form [2]

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \quad (3)$$

where $D_n = \max(D_n^+, D_n^-)$,

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i, \theta) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ F(x_i, \theta) - \frac{i-1}{n} \right\},$$

n is the sample size, x_1, x_2, \dots, x_n are sample values in increasing order is usually used. The distribution of statistic (3) obeys the Kolmogorov distribution law $K(S)$ [1] in testing simple hypotheses.

For verification of ω^2 Cramer-Von Mises-Smirnov goodness-of-fit test is used a statistic of the form

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \quad (4)$$

and in test of Ω^2 Anderson-Darling type [3, 4] the statistic of the form

$$S_\Omega = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n} \right) \ln(1 - F(x_i, \theta)) \right\}. \quad (5)$$

In verification a simple hypothesis statistic (4) obeys the $a1(S)$ distribution and statistic (5) obeys the $a2(S)$ distribution (see [1]).

In verification of composite hypotheses the conditional distribution law of the statistic $G(S|H_0)$ is affected by a number of factors: the form of the observed law $F(x, \theta)$ corresponding to the true hypothesis H_0 ; the type of the parameter estimation and the number of estimated parameters; sometimes it is a specific value of the parameter (e.g., in the case of gamma-distribution and beta-distribution families); the method of parameter estimation. The distinctions in the marginal distributions of the same statistics in testing simple and composite hypotheses are so significant that we cannot neglect them.

The paper [5] was one of the first in investigating statistic distributions of the nonparametric goodness-of-fit tests with composite hypotheses. Then, for the solution to this problem, various approaches were used [6, 7], [8-10], [11, 12], [13], [14], [15-16], [17], [18, 19], [20], [21, 22].

In our research [23-28] statistic distributions of the nonparametric goodness-of-fit tests are investigated by the methods of statistical simulating and for constructed empirical distributions approximate models of law were founded. Obtained results were used for developing of recommendations for standardization [29]. Precise models of statistics distributions and the tables of upper percentage points presented in [30-34]. The comparative analysis results of power of goodness-of-fit tests in verification a single and composite hypothesis are presented in the paper [35-37].

3 Distributions of statistics of the tests in the case of verification composite hypotheses concerning generalized Weibull distribution

In the case of verification composite hypotheses concerning Generalized Weibull Distribution the distributions of statistics $G(S|H_0)$ for nonparametric goodness-of-fit tests depend on specific values of shape parameters θ_0 .

In the fig. 1 you can see the behavior of statistics distribution S_Ω in testing composite hypotheses for family (1). In the case when two parameters are estimated by MLM (fig. 1) you can see the following: with the growing of values of parameters θ_1 the distribution $G(S|H_0)$ shifts to the right.

In the case when three parameters are estimated by MLM you can see the next: when the values of shape parameter are grow up till $\theta_1 \approx 2$ the distribution $G(S|H_0)$ is shift to the left. With the following growth of values of shape parameter the distribution $G(S|H_0)$ shifts to the opposite direction, to the right.

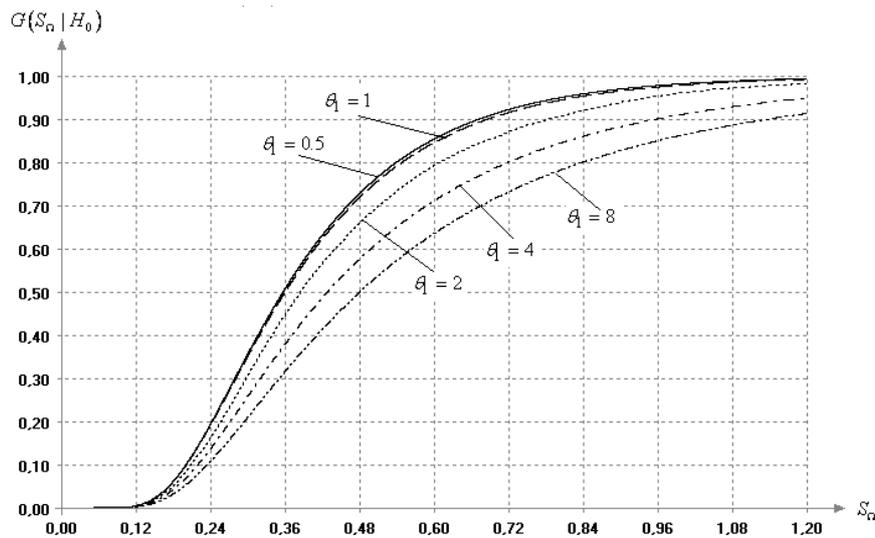


Fig. 1. Statistic distributions (5) of Anderson-Darling goodness-of-fit tests in composite hypotheses testing concerning family (1). MLM is used for estimation two (θ_0 and θ_1) parameters.

Percentage points obtained by statistic modeling and the models of marginal statistic distributions of Kolmogorov, Cramer-Von Mises-Smirnov and Anderson-Darling tests were computed for the values of shape parameter $\theta_1 = 1.0$ with MLM

used for parameter estimation are presented in table 1. Similar tables are constructed for the values of shape parameter $\theta_1 = 0.5, 2, 3, 4, 5, 6, 7, 8$.

Distributions $G(S|H_0)$ of the Kolmogorov, Cramer-Von Mises-Smirnov and the Anderson-Darling statistics are best approximated by the family of the III type beta-distributions with the density function

$$B_3(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = \frac{\theta_2^{\theta_0}}{\theta_3 B(\theta_0, \theta_1)} \frac{\left(\frac{x-\theta_4}{\theta_3}\right)^{\theta_0-1} \left(1-\frac{x-\theta_4}{\theta_3}\right)^{\theta_1-1}}{\left[1+(\theta_2-1)\frac{x-\theta_4}{\theta_3}\right]^{\theta_0+\theta_1}},$$

or by the family of the *Sb*-Johnson distributions

$$Sb(\theta_0, \theta_1, \theta_2, \theta_3) = \frac{\theta_1 \theta_2}{(x-\theta_3)(\theta_2+\theta_3-x)} \exp\left\{-\frac{1}{2}\left(\theta_0 - \theta_1 \ln \frac{x-\theta_3}{\theta_2+\theta_3-x}\right)^2\right\}.$$

The tables of percentage points and statistic distributions models were constructed by modeled statistic samples with the size $N = 10^6$. In the case when $N = 10^6$ the deviation the empirical p.d.f. $G_N(S|H_0)$ from the theoretical one is less than 10^{-3} . In this case the values of statistics of goodness-of-fit- tests were calculated using a samples of pseudorandom variables which belong to $F(x, \theta)$ with sample size $n = 10^3$. For the such value of n statistic p.d.f. $G(S_n|H_0)$ almost coincides with the marginal p.d.f. $G(S|H_0)$.

Table 1. Percentage points and models of limiting statistic distributions of the nonparametric goodness-of-fit test when MLM is used for parameter estimation (for $\theta_1 = 1$)

| Parameter estimated | Percentage points | | | Model |
|--------------------------------|-------------------|-------|-------|---|
| | 0.9 | 0.95 | 0.99 | |
| for Kolmogorov's test | | | | |
| θ_0 | 1.181 | 1.316 | 1.585 | $B_3(6.9734, 4.8247, 5.3213, 2.3800, 0.2690)$ |
| θ_1 | 1.083 | 1.196 | 1.425 | $B_3(4.6425, 6.6688, 2.8491, 2.2246, 0.3200)$ |
| θ_2 | 0.994 | 1.092 | 1.290 | $B_3(6.2635, 7.1481, 3.2059, 2.0000, 0.2800)$ |
| θ_0, θ_1 | 0.874 | 0.954 | 1.117 | $Sb(2.4299, 1.8866, 1.7504, 0.2598)$ |
| θ_0, θ_2 | 0.823 | 0.893 | 1.033 | $B_3(5.8989, 7.5040, 2.4180, 1.3724, 0.2800)$ |
| θ_1, θ_2 | 0.815 | 0.883 | 1.023 | $Sb(2.4499, 1.9720, 1.6016, 0.2486)$ |
| $\theta_0, \theta_1, \theta_2$ | 0.758 | 0.820 | 0.946 | $Sb(2.3012, 1.9386, 1.3863, 0.2464)$ |

| for Cramer-Von Mises-Smirnov's test | | | | |
|-------------------------------------|-------|-------|-------|--|
| θ_0 | 0.320 | 0.431 | 0.706 | B_3 (2.2422, 2.2970, 16.4663, 1.6500, 0.0130) |
| θ_1 | 0.227 | 0.295 | 0.464 | B_3 (5.3830, 2.6954, 40.5199, 1.6450, 0.0050) |
| θ_2 | 0.174 | 0.221 | 0.336 | B_3 (3.6505, 3.2499, 16.5445, 1.0000, 0.0100) |
| θ_0, θ_1 | 0.117 | 0.144 | 0.209 | Sb (3.8667, 1.4603, 0.7583, 0.0059) |
| θ_0, θ_2 | 0.102 | 0.123 | 0.174 | B_3 (12.2776, 4.1107, 27.2069, 0.4875, 0.0000) |
| θ_1, θ_2 | 0.103 | 0.127 | 0.182 | B_3 (4.7144, 4.6690, 10.8816, 0.5261, 0.0059) |
| $\theta_0, \theta_1, \theta_2$ | 0.080 | 0.097 | 0.135 | Sb (4.1842, 1.6587, 0.4794, 0.0061) |
| for Anderson-Darling's test | | | | |
| θ_0 | 1.724 | 2.280 | 3.639 | B_3 (4.8106, 2.6855, 35.5593, 11.8700, 0.0500) |
| θ_1 | 1.275 | 1.617 | 2.468 | B_3 (3.6999, 3.9108, 16.4841, 9.0300, 0.0740) |
| θ_2 | 1.056 | 1.314 | 1.953 | B_3 (4.9871, 4.1479, 16.5432, 6.4500, 0.0600) |
| θ_0, θ_1 | 0.687 | 0.827 | 1.161 | B_3 (4.6368, 6.6727, 7.1680, 3.6356, 0.0521) |
| θ_0, θ_2 | 0.633 | 0.753 | 1.037 | B_3 (3.0467, 5.9239, 5.0944, 2.7870, 0.1000) |
| θ_1, θ_2 | 0.696 | 0.842 | 1.194 | B_3 (6.9638, 4.5238, 17.7792, 3.8000, 0.0522) |
| $\theta_0, \theta_1, \theta_2$ | 0.494 | 0.582 | 0.786 | Sb (3.9578, 1.6861, 2.5760, 0.0547) |

4 Conclusions

The Generalized Weibull probability distribution plays an important role in a statistical analysis of lifetime or response data in reliability and survival studies. In certain parameters the function of Generalized Weibull Distribution agree with the Weibull distribution function and the exponential distribution function. In this work you can find how density of distribution depends on the parameters values of the law.

In this work are presented models of the statistic distributions of the nonparametric goodness-of-fit tests for testing composite hypotheses with the distributions family (1).

It should be stressed, that obtained percentage points and models guarantee proper implementation of the nonparametric goodness-of-fit tests in statistic analysis problems if MLM is used. These results can't be used with other estimations because statistic distributions of these tests are essential depend on estimation method [25].

The authors hope that release of the article will be conducive to decrease mistake amount, committed in statistic analysis problems if nonparametric goodness-of-fit tests are used [27].

Acknowledgments

This research was supported by the Russian Foundation for Basic Research (project № 09-01-00056a) and by Federal Education Agency of Russian Federation Ministry of Education and Science in the Analytical departmental purposeful program framework "Development of Higher School Potential" and in the framework of Federal task program "Scientific and scientific-educational staff of the innovation Russia".

References

1. Bolshev L.N., On the question on testing some composite statistical hypotheses. In *Theory of Probability and Mathematical Statistics*. Selected Works. Nauka, Moscow, 5-63 (1987).
2. Bolshev L.N., Smirnov N.V., Tables of Mathematical Statistics. Moscow: Science (1983). (in Russian)
3. Anderson T.W., and Darling D.A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Ann. Math. Statist.*, 23, 193-212 (1952).
4. Anderson T.W., and Darling D.A., A test of goodness of fit, *J. Amer. Statist. Assoc.*, 29, 765-769 (1954).
5. Kac M., Kiefer J., Wolfowitz J., On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods. *Ann. Math. Stat.*, 26, 189-211 (1955).
6. Darling D.A. The Cramer-Smirnov test in the parametric case, *Ann. Math. Statist.*, 26, 1-20 (1955).
7. Darling D.A. The Cramer-Smirnov test in the parametric case, *Ann. Math. Statist.*, 28, 823-838 (1957).
8. Durbin J., Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.*, 1, 279-290 (1973).
9. Durbin J., Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests of spacings, *Biometrika*, 62, 5-22 (1975).
10. Durbin J., Kolmogorov-Smirnov Test when Parameters are Estimated, *Lect. Notes Math.*, 566, 33-44 (1976).
11. Gihman I.I., Some remarks on the consistency criterion of A.N. Kolmogorov. *Dokl. Akad. Nauk SSSR*, 91, 4, 715-718 (1953).
12. Gihman I.I., On the empirical distribution function in the case of grouping data, In *Selected Translations in Mathematical Statistics and Probability*, 1, American Math. Soc., Providence, RI, 77-81, (1961).
13. Martinov G.V., Omega-Quadrat Tests. *Moscow: Science* (1978). (in Russian)

14. Pearson E.S., Hartley H.O. *Biometrika Tables for Statistics*. Vol. 2. Cambridge: University Press (1972).
15. Stephens M.A., Use of Kolmogorov–Smirnov, Cramer – von Mises and Related Statistics – Without Extensive Table, *J. R. Stat. Soc.*, 32, 115-122 (1970).
16. Stephens M.A., EDF Statistics for Goodness of Fit and Some Comparisons, *J. Am. Statist. Assoc.*, 69, 730-737 (1974).
17. Chandra M., Singpurwalla N.D., Stephens M.A., Statistics for Test of Fit for the Extrem–Value and Weibull Distribution, *J. Am. Statist. Assoc.*, 76, 375, 729-731 (1981).
18. Tyurin Yu.N., On the Limiting Kolmogorov-Smirnov Statistic Distribution for Composite Hypothesis, *News of the AS USSR, Ser. Math.*, 48, 6, 1314-1343 (1984). (in Russian)
19. Tyurin Yu.N., Savvushkina N.E., Goodness-of-Fit Test for Weibull-Gnedenko Distribution, *News of the AS USSR, Ser. Techn. Cybernetics*, 3, 109-112 (1984). (in Russian)
20. Dzhaparidze K.O., and Nikulin M.S., Probability distribution of the Kolmogorov and omega-square statistics for continuous distributions with shift and scale parameters, *J. Soviet Math.*, 20, 2147-2163 (1982).
21. Nikulin M.S., Gihman and goodness-of-fit tests for grouped data, *Mathematical Reports of the academy of Science of the Royal Society of Canada*, 14, 4, 151-156 (1992).
22. Nikulin M.S., A variant of the generalized omega-square statistic, *J. Soviet Math.*, 61, 4, 1896-1900 (1992).
23. Lemeshko B.Yu., Postovalov S.N., Statistical Distributions of Nonparametric Goodness-of-Fit Tests as Estimated by the Sample Parameters of Experimentally Observed Laws, *Industrial laboratory*, 64, 3, 197-208 (1998).
24. Lemeshko B.Yu., Postovalov S.N., Application of the Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses. *Optoelectronics, Instrumentation and Data Processing*, 37, 2, 76-88 (2001).
25. Lemeshko B.Yu., Postovalov S.N. About dependence of statistics distributions of the nonparametric tests and their power from a method of the parameters estimation, *Industrial laboratory. Diagnostics of materials*, 6, 77, 62-71 (2001). (in Russian)
26. Lemeshko B.Yu., Postovalov S.N., The nonparametric goodness-of-fit tests about fit with Johnson distributions in testing composite hypotheses. *News of the SB AS HS*, 1, 5, 65-74 (2002). (in Russian)
27. Lemeshko B.Yu., Errors when using nonparametric fitting criteria, *Measurement Techniques*, 47, 2, – P.134-142 (2004).
28. Lemeshko B.Yu., Maklakov A.A., Nonparametric Test in Testing Composite Hypotheses on Goodness of Fit Exponential Family Distributions, *Optoelectronics, Instrumentation and Data Processing*, 40, 3, 3-18 (2004).

29. R 50.1.037-2002. Recommendations for Standardization. Applied statistics. Rules of check of experimental and theoretical distribution of the consent. Part II. Nonparametric goodness-of-fit test. *Moscow: Publishing house of the standards*, (2002). (in Russian)
30. Lemeshko S.B., Lemeshko B.Yu., Statistic distributions of the nonparametric goodness-of-fit tests in testing hypotheses relative to beta-distributions, *News of the SB AS HS*, 2, 9, 6-16 (2007). (in Russian)
31. Lemeshko B.Yu., Lemeshko S.B., Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Part 1, *Measurement Techniques*, 52, 6, 555-565 (2009).
32. Lemeshko B.Yu., Lemeshko S.B., Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Part II. *Measurement Techniques*, 52, 8, 799-812 (2009).
33. Lemeshko B.Yu., Lemeshko S.B., Models of Statistic Distributions of Nonparametric Goodness-of-fit Tests in Composite Hypotheses Testing in Case of Double Exponential Law, *The XIII International Conference "Applied Stochastic Models and Data Analysis"* (ASMDA-2009), June 30-July 3, Selected papers. Vilnius, Lithuania, 153-157 (2009).
34. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Statistic Distribution Models for some nonparametric goodness-of-fit tests in testing composite hypothesis, *Communications in Statistics - Theory and Methods*, 39, 3, 460-471, (2010).
35. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., The power of goodness-of-fit tests for close alternatives, *Measurement Techniques*, 50, 2, 132-141 (2007).
36. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses, *Journal of Applied and Industrial Mathematics*, 3, 4, 462-475 (2009).
37. Lemeshko B.Yu. Lemeshko S.B., Postovalov S.N., Comparative analysis of the power of goodness-of-fit tests for near competing hypotheses. II. Verification of complex hypotheses, *Journal of Applied and Industrial Mathematics*, 4, 1, 79-93 (2010).

Tests For Homogeneity Of Variances Under Violation Of Normality Assumption

Boris Yu. Lemeshko, Stanislav B. Lemeshko and Alice A. Gorbunova

Novosibirsk State Technical University, Department of applied mathematics, Novosibirsk, Russia,
e-mail: Lemeshko@fpm.ami.nstu.ru

Abstract. The comparative analysis of power of classical variance homogeneity tests (Fisher's, Bartlett's, Cochran's, Hartley's and Levene's tests) is carried out. Distributions of tests statistics are investigated under violation of assumptions that samples belong to the normal law. Distributions and power of nonparametric tests for dispersion characteristics homogeneity are researched (Ansari-Bradley's, Mood's, Siegel-Tukey's tests). The comparative analysis of power of classical variance homogeneity tests with power of nonparametric tests is carried out. Tables of percentage points for Cochran's test are presented in case of the distributions which are different from normal.

Key words: test of variances homogeneity, Fisher's test, Bartlett's test, Cochran's test, Hartley's test, Levene's test, nonparametric test, Ansari-Bradley's test, Mood's test, Siegel-Tukey's test, power of test.

1 Introduction

Tests of samples homogeneity are often used in various applications of statistical analysis. The question can be about checking hypotheses about homogeneity of samples distributions, population means or variances. Naturally the most complete findings can be done in the first case. However researcher can be interested in possible deviations in the sample mean values or differences in dispersion characteristics of measurements results.

Application features of nonparametric Smirnov and Lehmann-Rosenblatt homogeneity tests and analysis of their power were considered in [1]. In [2] it was shown that classical criteria for testing hypotheses about homogeneity of means are stable to violation of normality assumption and comparative analysis of the power of various tests, including nonparametric, was given.

One of the basic assumptions in constructing classical tests for equality of variances is normal distribution of observable random variables (measurement errors). Therefore the application of classical criteria always involves the question of how valid the results obtained are in this particular situation. Under violation of assumption that analyzed variables belong to normal law, conditional distributions of tests statistics, when hypothesis under test is true, change appreciably.

All available publications do not give full information on the power of the classical tests for homogeneity of variances and on comparative analysis of the power of the

classical tests and nonparametric criteria for testing hypotheses about the homogeneity of the dispersion characteristics (scale parameters).

This work continues researches of stability of criteria for testing hypotheses about the equality of variances [3]. Classical Bartlett's [4], Cochran's [5], Fisher's, Hartley's [6], Levene's [7] tests have been compared, nonparametric (rank) Ansari-Bradley's [8], Mood's [9], Siegel-Tukey's [10] tests have been considered. The purpose of the paper is

- research of statistics distributions for listed tests in case of distribution laws of observable random variables which are different from normal;
- comparative analysis of criteria power concerning concrete competing hypotheses;
- realization of the possibility to apply the classical tests under violation of assumptions about normality of random variables.

A hypothesis under test for equality of variances corresponding to m samples will have the form

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2, \quad (1)$$

and the competitive hypothesis is

$$H_1 : \sigma_{i_1}^2 \neq \sigma_{i_2}^2, \quad (2)$$

where the inequality holds at least for one pair of subscripts i_1, i_2 .

Statistical simulation methods and the developed software have been used for investigating statistic distributions, calculating percentage points and estimating tests power with respect to various competing hypotheses. The sample size of statistics under study was $N = 10^6$. Such N allowed absolute value of difference between true law of statistics distribution and simulated empirical not to exceed 10^{-3} .

Statistic distributions have been studied for various distribution laws, in particular, in case when simulated samples belong to the family with density

$$De(\theta_0) = f(x; \theta_0, \theta_1, \theta_2) = \frac{\theta_0}{2\theta_1\Gamma(1/\theta_0)} \exp\left(-\left(\frac{|x-\theta_2|}{\theta_1}\right)^{\theta_0}\right) \quad (3)$$

with various values of the form parameter θ_0 . This family can be a good model for error distributions of various measuring systems. Special cases of distribution $De(\theta_0)$ include the Laplace ($\theta_0 = 1$) and normal ($\theta_0 = 2$) distribution. The family (3) allows to define various symmetric distributions that differ from normal: the smaller value of form parameter θ_0 the "heavier" tails of the distribution $De(\theta_0)$, and vice-versa the higher value the "easier" tails.

The competing hypotheses of the form $H_1: \sigma_m = d\sigma_0$ have been considered in comparative analysis of the test power. That is, a competing hypothesis corresponds to the situation when $m-1$ samples belong to the law with $\sigma = \sigma_0$, while one of the samples, for example, with number m has some different variance. Hypothesis under test corresponds to the situation $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2 = \sigma_0^2$.

2 Bartlett's test

Bartlett's test statistic [4] is

$$\chi^2 = M \left[1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \frac{1}{v_i} - \frac{1}{N} \right) \right]^{-1}, \quad (4)$$

where

$$M = N \ln \left(\frac{1}{N} \sum_{i=1}^m v_i S_i^2 \right) - \sum_{i=1}^m v_i \ln S_i^2,$$

m is the number of samples; n_i are the sample sizes; $v_i = n_i$, if mathematical expectation is known, and $v_i = n_i - 1$, if it is unknown; $N = \sum_{i=1}^m v_i$; S_i^2 – estimators of the sample variances. If the mathematical expectation is unknown, the estimators are $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ji} - \bar{X}_i)^2$, where X_{ji} – j -th observation in sample i , $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ji}$.

If hypothesis H_0 is true, all $v_i > 3$ and samples are extracted from a normal population, then the statistic (4) has approximately the χ_{m-1}^2 distribution. If measurements are normally distributed, the distribution for the statistic (4) is almost independent of the sample sizes n_i [3]. If distributions of observed variables differ from the normal law, the distribution $G(\chi^2 | H_0)$ of statistic (4) becomes depending on n_i and differs from χ_{m-1}^2 .

3 Cochran's test

When all n_i are equal, one can use simpler Cochran's test [5]. The test statistic Q is defined as follows:

$$Q = \frac{S_{\max}^2}{S_1^2 + S_2^2 + \dots + S_m^2}, \quad (5)$$

where $S_{\max}^2 = \max(S_1^2, S_2^2, \dots, S_m^2)$, m is the number of independent estimators of variances (number of samples), S_i^2 are estimators of the sample variances.

Distribution of Cochran's test statistic strongly depends on the sample size. The reference literature gives only tables of the percentage points for limited number of values n , which are used in hypothesis testing.

4 Hartley's test

Hartley's test [6] as well as Cochran's test is used in case of samples of equal size. Hartley's test statistic for homogeneity of variances is

$$F = \frac{S_{\max}^2}{S_{\min}^2}, \quad (6)$$

where $S_{\max}^2 = \max(S_1^2, S_2^2, \dots, S_m^2)$, $S_{\min}^2 = \min(S_1^2, S_2^2, \dots, S_m^2)$, m - number of independent estimators of variances (number of samples).

Literature gives tables of percentage points for distribution of statistic (6) depending on $v_1 = m$ and $v_2 = n - 1$.

5 Levene's test

The Levene's test statistic [7] is defined as:

$$W = \frac{N - m \sum_{i=1}^m n_i (\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2}{m - 1 \sum_{i=1}^m \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i\cdot})^2}, \quad (7)$$

where m is the number of samples, n_i is the sample size of the i -th sample,

$N = \sum_{i=1}^m n_i$, $Z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$, X_{ij} - j -th observation in sample i , $\bar{X}_{i\cdot}$ is the mean

of i -th sample, $\bar{Z}_{i\cdot}$ is the mean of the Z_{ij} for sample i , $\bar{Z}_{\cdot\cdot}$ - the mean of all Z_{ij} .

In some descriptions of the test, for example [11], it is said that in case when samples belong to the normal law and hypothesis H_0 is true, the statistic has a F_{v_1, v_2} -distribution with number of degrees of freedom $v_1 = m - 1$ and $v_2 = N - m$.

Actually *distribution of statistics (7) is not Fisher's distribution F_{v_1, v_2}* . Therefore

percentage points of distribution were investigated using statistical simulation methods [12].

Levene's test is less sensitive to departures from normality. However it has less power.

The original Levene's test used only sample means. Brown and Forsythe [13] suggested using sample median and trimmed mean as estimators of the mean for statistic (7).

However our researches have shown that *using in (7) sample median and trimmed mean leads to another distribution $G(W|H_0)$ of statistics (7).*

6 Fisher's test

Fisher's test is used to check hypothesis of variances homogeneity for *two* samples of random variables. The test statistic has a simple form

$$F = \frac{s_1^2}{s_2^2}, \quad (8)$$

where s_1^2 and s_2^2 – unbiased variance estimators, computed from the sample data.

In case when samples belong to the normal law and hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ is true, this statistic has the F_{v_1, v_2} -distribution with number of degrees of freedom

$v_1 = n_1 - 1$ and $v_2 = n_2 - 1$. A hypothesis under test is rejected if $F^* < F_{\alpha/2, v_1, v_2}$ or $F^* > F_{1-\alpha/2, v_1, v_2}$.

7 Comparative analysis of power

At given probability of type I error α (to reject the null hypothesis when it is true) it is possible to judge advantages of the test by value of power $1 - \beta$, where β is the probability of type II error (not to reject the null hypothesis when alternative is true). In [14] it is definitely said that Cochran's test has lower power in comparison with Bartlett's test. In [3] it was shown that Cochran's test has greater power by the example of checking hypothesis about variances homogeneity for *five* samples.

Research of power of Bartlett's, Cochran's, Hartley's, Fisher's and Levene's tests concerning such competing hypotheses $H_1: \sigma_2 = d\sigma_1$, $d \neq 1$ (in case of two samples that belong to the normal law) has shown that Bartlett's, Cochran's, Hartley's and Fisher's tests have equal power in this case. Levene's test appreciably yields to them in power.

In case of the distributions which are different from normal, for example, family of distributions with density (3), Bartlett's, Cochran's, Hartley's and Fisher's tests

remain equivalent in power, and Levene's test also appreciably yields to them. However in case of heavy-tailed distributions (for example, when samples belong to the Laplace distribution) Levene's test has advantage of greater power. Bartlett's, Cochran's, Hartley's and Levene's tests can be applied when number of samples $m > 2$. In such situations power of these tests is different. If $m > 2$ and normality assumption is true, given tests can be ordered by power decrease as follows:

$$\text{Cochran's } \succ \text{Bartlett's } \succ \text{Hartley's } \succ \text{Levene's.}$$

The preference order remains in case of violation of normality assumption. The exception concerns situations when samples belong to laws with more "heavy tails" in comparison with the normal law. For example, in case of Laplace distribution Levene's test is more powerful than three others.

8 Ansari-Bradley's test

Nonparametric analogues of tests for homogeneity of variances are used to check hypothesis that two samples with sample sizes n_1 and n_2 belong to population with identical characteristics of dispersion. As a rule equality of means is supposed.

The Ansari-Bradley's test statistic [8] is:

$$S = \sum_{i=1}^{n_1} \left\{ \frac{n_1 + n_2 + 1}{2} - \left| R_i - \frac{n_1 + n_2 + 1}{2} \right| \right\}, \quad (9)$$

where R_i - ranks corresponding to elements of the first sample in general variational row. In case when samples belong to the same law and checked hypothesis H_0 is true, distribution of statistics (9) does not depend on this law. *Discreteness* of distribution of statistics (9) can be practically neglected when $n_1, n_2 > 40$.

9 Siegel-Tukey's test

The variational row constructed on general sample $x_1 \leq x_2 \leq \dots \leq x_n$, where $n = n_1 + n_2$, is transformed into such sequence

$$x_1, x_n, x_{n-1}, x_2, x_3, x_{n-2}, x_{n-3}, x_4, x_5, \dots,$$

i.e. row of remained values is "turned over" each time when ranks are assigned to pair of extreme values. Sum of ranks of sample with smaller size is used as test statistics. When $n_1 < n_2$ test statistic is defined as:

$$R = \sum_{i=1}^{n_1} R_i. \quad (10)$$

Discreteness of distribution of statistics (10) can be practically neglected when $n_1, n_2 > 30$.

10 Mood's test

The test statistic is [9]:

$$M = \sum_{i=1}^{n_1} \left(R_i - \frac{n_1 + n_2 + 1}{2} \right)^2, \quad (11)$$

where R_i - ranks of sample with smaller size in general variational row. *Discreteness* of distribution of statistics (11) can be neglected at all when $n_1, n_2 > 20$.

When sample sizes $n_1, n_2 > 10$ discrete distributions of statistics (9), (10) and (11) are well enough approximated by normal law. Therefore instead of statistics (9), (10) and (11) normalized analogues are more often used, which are approximately standard normal.

Results of power research have shown appreciable advantage of Mood's test and practical equivalence of Siegel-Tukey's and Ansari-Bradley's tests. Of course, nonparametric tests yield in power to Bartlett's, Cochran's, Hartley's and Fisher's tests. Figure 1 shows graphs of criteria power concerning competing hypotheses $H_1^1: \sigma_2 = 1.1\sigma_1$ and $H_1^2: \sigma_2 = 1.5\sigma_1$ depending on sample size n_i in case when $\alpha = 0.1$ and samples belong to the normal law. As we see, advantage in power of Cochran's test is rather significant in comparison with Mood's test - most powerful of nonparametric tests. Let's remind that Bartlett's, Cochran's, Hartley's and Fisher's tests have equal power in case of two samples.

Distributions of nonparametric tests statistics do not depend on a law kind, if both samples belong to the same population. But if samples belong to different laws and hypothesis of variances equality H_0 is true, *distributions of statistics of nonparametric tests depend on a kind of these laws*.

11 Cochran's test in case of laws different from normal

Classical tests have considerable advantage in power over nonparametric. This advantage remains when analyzed samples belong to the laws appreciably different from normal. Therefore there is every reason to research statistics distributions of classical tests for checking variances homogeneity (construction of distributions models or tables of percentage points) in case of laws most often used in practice (different from the normal law). Among considered tests Cochran's test is the most suitable for this role.

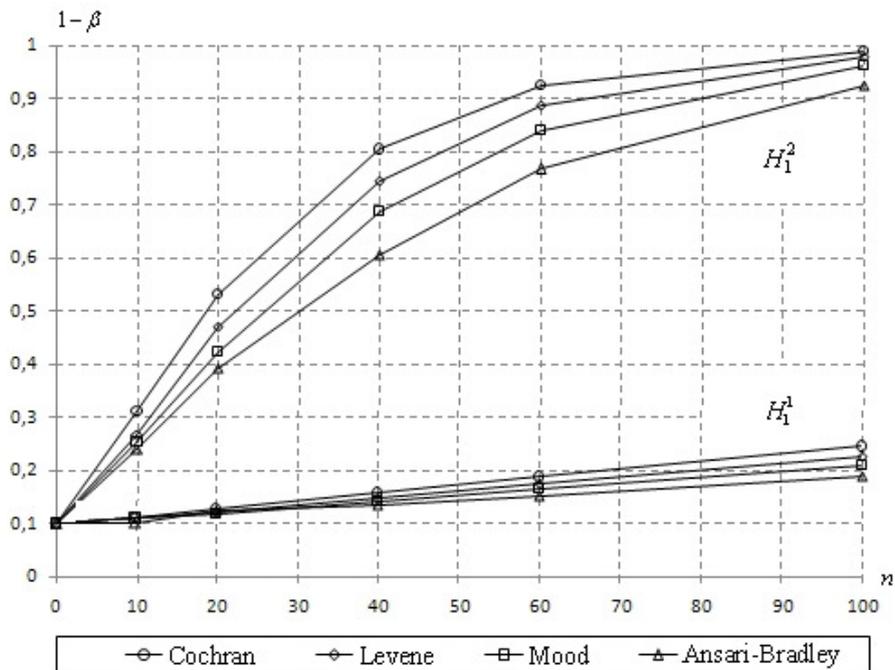


Fig. 1. Power of tests concerning competing hypotheses H_1^1 and H_1^2 depending on sample size n when $\alpha = 0.1$ and samples belong to normal law

In case when observable variables belong to family of distributions (3) with parameter of the form $\theta_0 = 1, 2, 3, 4, 5$ and some values n , tables of upper percentage points (1%, 5%, 10%) for Cochran's test were obtained using statistical simulation (when number of samples $m = 2 \div 5$). The results obtained can be used in situations when distribution (3) with appropriate parameter θ_0 is a good model for observable random variables. Computed percentage points improve some results presented in [3] and expand possibilities to apply Cochran's test.

12 Acknowledgments

This research was supported by the Russian Foundation for Basic Research (project no. 09-01-00056a), by the Federal Agency for Education within the framework of the analytical domestic target program "Development of the scientific potential of higher schools" and federal target program of the Ministry of Education and Science of the Russian Federation "Scientific and scientific-pedagogical personnel of innovative Russia".

References

1. Lemeshko B.Yu., Lemeshko S.B., Statistical distribution convergence and homogeneity test power for Smirnov and Lehmann–Rosenblatt tests, *Measurement Techniques*, **48**, 12, 1159–1166 (2005).
2. Lemeshko B.Yu., Lemeshko S.B., Power and robustness of criteria used to verify the homogeneity of means, *Measurement Techniques*, **51**, 9, 950–959 (2008).
3. Lemeshko B.Yu., Mirkin E.P., Bartlett and Cochran tests in measurements with probability laws different from normal, *Measurement Techniques*, **47**, 10, 960–968 (2004).
4. Bartlett M.S., Properties of sufficiency of statistical tests, *Proc. Roy. Soc. A* **160**, 268–287 (1937).
5. Cochran W.G., The distribution of the largest of a set of estimated variances as a fraction of their total, *Annals of Eugenics*, **11**, 47–52 (1941).
6. Hartley H.O., The maximum F-ratio as a short-cut test of heterogeneity of variance, *Biometrika*, **37**, 308–312 (1950).
7. Levene H., Robust tests for equality of variances, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, 278–292 (1960).
8. Ansari A.R., Bradley R.A., Rank-tests for dispersions, *AMS*, **31**, 4, 1174–1189 (1960).
9. Mood A., On the asymptotic efficiency of certain nonparametric tests, *AMS*, **25**, 514–522 (1954).
10. Siegel S., Tukey J.W., A nonparametric sum of rank procedure for relative spread in unpaired samples, *JASA*, **55**, 91, 429–445 (1960).
11. Levene Test for Equality of Variances. In: e-Handbook of Statistical Methods (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>).
12. Neel J.H., Stallings W.M., A Monte Carlo Study of Levene's Test of Homogeneity of Variance: Empirical Frequencies of Type I Error in Normal Distributions. (Paper presented at the Annual Meeting of the American Educational Research Association Convention (Chicago, Illinois, April 1974).
13. Brown M.B., Forsythe A.B., Robust Tests for Equality of Variances, *J. Amer. Statist. Assoc.*, **69**, 364–367 (1974).
14. Bol'shev L.N., Smirnov N.V., *Tables of Mathematical Statistics*. Nauka, Moscow (1983). [in Russian].

APPENDIX

Table 1. Upper percentage points for Cochran's test statistic distribution in case of two samples with equal size n

| n | $De(1)$ | | | $De(2)$ | | | $De(3)$ | | | $De(4)$ | | | $De(5)$ | | |
|-----|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|
| | α | | | α | | | α | | | α | | | α | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 5 | 0.917 | 0.947 | 0.980 | 0.865 | 0.906 | 0.959 | 0.845 | 0.890 | 0.950 | 0.836 | 0.883 | 0.947 | 0.831 | 0.879 | 0.945 |
| 8 | 0.862 | 0.900 | 0.949 | 0.791 | 0.833 | 0.899 | 0.764 | 0.807 | 0.877 | 0.751 | 0.794 | 0.866 | 0.744 | 0.787 | 0.861 |
| 10 | 0.836 | 0.875 | 0.930 | 0.761 | 0.801 | 0.868 | 0.733 | 0.773 | 0.842 | 0.720 | 0.759 | 0.829 | 0.713 | 0.751 | 0.822 |
| 15 | 0.789 | 0.829 | 0.890 | 0.713 | 0.748 | 0.811 | 0.686 | 0.719 | 0.780 | 0.674 | 0.706 | 0.765 | 0.667 | 0.698 | 0.757 |
| 20 | 0.759 | 0.797 | 0.858 | 0.684 | 0.716 | 0.774 | 0.660 | 0.689 | 0.743 | 0.648 | 0.676 | 0.728 | 0.642 | 0.669 | 0.720 |
| 25 | 0.736 | 0.772 | 0.834 | 0.665 | 0.694 | 0.748 | 0.642 | 0.668 | 0.717 | 0.632 | 0.656 | 0.703 | 0.626 | 0.649 | 0.695 |
| 30 | 0.718 | 0.753 | 0.814 | 0.650 | 0.677 | 0.727 | 0.629 | 0.653 | 0.699 | 0.619 | 0.642 | 0.685 | 0.614 | 0.635 | 0.677 |
| 40 | 0.693 | 0.725 | 0.782 | 0.630 | 0.654 | 0.699 | 0.611 | 0.632 | 0.672 | 0.603 | 0.622 | 0.660 | 0.598 | 0.616 | 0.653 |
| 50 | 0.674 | 0.704 | 0.758 | 0.617 | 0.638 | 0.679 | 0.599 | 0.618 | 0.654 | 0.591 | 0.609 | 0.642 | 0.587 | 0.604 | 0.636 |
| 60 | 0.660 | 0.689 | 0.740 | 0.606 | 0.626 | 0.664 | 0.591 | 0.608 | 0.640 | 0.583 | 0.599 | 0.630 | 0.579 | 0.594 | 0.624 |
| 70 | 0.649 | 0.676 | 0.724 | 0.598 | 0.617 | 0.652 | 0.584 | 0.599 | 0.630 | 0.577 | 0.591 | 0.620 | 0.573 | 0.587 | 0.614 |
| 80 | 0.640 | 0.665 | 0.712 | 0.592 | 0.609 | 0.642 | 0.578 | 0.593 | 0.621 | 0.572 | 0.585 | 0.612 | 0.568 | 0.581 | 0.607 |

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
 International Conference, Chania, Crete, Greece, 8 - 11 June 2010

| | | | | | | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 90 | 0.632 | 0.657 | 0.701 | 0.587 | 0.603 | 0.634 | 0.573 | 0.587 | 0.614 | 0.567 | 0.580 | 0.605 | 0.564 | 0.576 | 0.600 |
| 100 | 0.626 | 0.649 | 0.692 | 0.582 | 0.598 | 0.628 | 0.570 | 0.583 | 0.609 | 0.564 | 0.576 | 0.600 | 0.561 | 0.572 | 0.595 |

Table 2. Upper percentage points for Cochran's test statistic distribution in case of three samples with equal size n

| n | $De(1)$ | | | $De(2)$ | | | $De(3)$ | | | $De(4)$ | | | $De(5)$ | | |
|-----|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|
| | α | | | α | | | α | | | α | | | α | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 5 | 0.794 | 0.847 | 0.918 | 0.700 | 0.752 | 0.839 | 0.665 | 0.717 | 0.806 | 0.649 | 0.700 | 0.790 | 0.641 | 0.690 | 0.781 |
| 8 | 0.716 | 0.768 | 0.852 | 0.614 | 0.658 | 0.741 | 0.579 | 0.620 | 0.698 | 0.563 | 0.602 | 0.677 | 0.554 | 0.591 | 0.665 |
| 10 | 0.681 | 0.732 | 0.817 | 0.581 | 0.622 | 0.698 | 0.548 | 0.584 | 0.654 | 0.533 | 0.567 | 0.634 | 0.524 | 0.557 | 0.622 |
| 15 | 0.623 | 0.669 | 0.751 | 0.531 | 0.564 | 0.628 | 0.503 | 0.531 | 0.588 | 0.489 | 0.516 | 0.569 | 0.482 | 0.508 | 0.558 |
| 20 | 0.587 | 0.629 | 0.707 | 0.502 | 0.531 | 0.588 | 0.477 | 0.501 | 0.550 | 0.466 | 0.488 | 0.533 | 0.459 | 0.480 | 0.524 |
| 25 | 0.562 | 0.600 | 0.673 | 0.484 | 0.509 | 0.560 | 0.461 | 0.482 | 0.526 | 0.450 | 0.470 | 0.510 | 0.444 | 0.463 | 0.501 |
| 30 | 0.543 | 0.578 | 0.647 | 0.470 | 0.493 | 0.539 | 0.449 | 0.468 | 0.507 | 0.439 | 0.457 | 0.493 | 0.434 | 0.451 | 0.485 |
| 40 | 0.515 | 0.547 | 0.608 | 0.450 | 0.470 | 0.510 | 0.432 | 0.449 | 0.482 | 0.424 | 0.439 | 0.470 | 0.419 | 0.434 | 0.463 |
| 50 | 0.496 | 0.525 | 0.581 | 0.437 | 0.455 | 0.490 | 0.421 | 0.436 | 0.465 | 0.414 | 0.427 | 0.454 | 0.410 | 0.422 | 0.448 |
| 60 | 0.482 | 0.508 | 0.560 | 0.428 | 0.444 | 0.476 | 0.413 | 0.426 | 0.453 | 0.406 | 0.418 | 0.443 | 0.402 | 0.414 | 0.437 |
| 70 | 0.471 | 0.495 | 0.543 | 0.421 | 0.435 | 0.465 | 0.407 | 0.419 | 0.444 | 0.401 | 0.412 | 0.434 | 0.397 | 0.408 | 0.429 |
| 80 | 0.462 | 0.485 | 0.530 | 0.415 | 0.429 | 0.456 | 0.402 | 0.413 | 0.436 | 0.396 | 0.406 | 0.427 | 0.393 | 0.403 | 0.422 |
| 90 | 0.455 | 0.476 | 0.518 | 0.410 | 0.423 | 0.449 | 0.398 | 0.408 | 0.430 | 0.392 | 0.402 | 0.422 | 0.389 | 0.398 | 0.417 |

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
 International Conference, Chania, Crete, Greece, 8 - 11 June 2010

| | | | | | | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 100 | 0.449 | 0.469 | 0.509 | 0.406 | 0.418 | 0.443 | 0.394 | 0.405 | 0.425 | 0.389 | 0.398 | 0.417 | 0.386 | 0.395 | 0.413 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|

Table 3. Upper percentage points for Cochran's test statistic distribution in case of four samples with equal size n

| n | $De(1)$ | | | $De(2)$ | | | $De(3)$ | | | $De(4)$ | | | $De(5)$ | | |
|-----|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|
| | α | | | α | | | α | | | α | | | α | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 5 | 0.696 | 0.755 | 0.848 | 0.584 | 0.634 | 0.727 | 0.545 | 0.591 | 0.679 | 0.527 | 0.571 | 0.656 | 0.517 | 0.560 | 0.643 |
| 8 | 0.611 | 0.666 | 0.761 | 0.501 | 0.541 | 0.619 | 0.466 | 0.500 | 0.569 | 0.450 | 0.482 | 0.546 | 0.441 | 0.471 | 0.533 |
| 10 | 0.575 | 0.626 | 0.720 | 0.470 | 0.506 | 0.576 | 0.438 | 0.468 | 0.529 | 0.423 | 0.451 | 0.507 | 0.415 | 0.441 | 0.495 |
| 15 | 0.517 | 0.561 | 0.646 | 0.424 | 0.453 | 0.510 | 0.397 | 0.421 | 0.468 | 0.385 | 0.406 | 0.450 | 0.378 | 0.398 | 0.439 |
| 20 | 0.482 | 0.521 | 0.598 | 0.399 | 0.422 | 0.471 | 0.375 | 0.395 | 0.435 | 0.364 | 0.382 | 0.419 | 0.358 | 0.375 | 0.410 |
| 25 | 0.457 | 0.493 | 0.563 | 0.382 | 0.403 | 0.445 | 0.360 | 0.378 | 0.413 | 0.351 | 0.366 | 0.398 | 0.346 | 0.360 | 0.390 |
| 30 | 0.439 | 0.471 | 0.536 | 0.369 | 0.388 | 0.427 | 0.350 | 0.365 | 0.397 | 0.341 | 0.355 | 0.384 | 0.336 | 0.349 | 0.377 |
| 40 | 0.413 | 0.441 | 0.498 | 0.352 | 0.368 | 0.401 | 0.335 | 0.348 | 0.348 | 0.328 | 0.340 | 0.364 | 0.324 | 0.335 | 0.358 |
| 50 | 0.395 | 0.420 | 0.470 | 0.340 | 0.355 | 0.384 | 0.326 | 0.337 | 0.361 | 0.319 | 0.329 | 0.351 | 0.315 | 0.325 | 0.345 |
| 60 | 0.382 | 0.404 | 0.451 | 0.332 | 0.345 | 0.371 | 0.319 | 0.329 | 0.350 | 0.313 | 0.322 | 0.341 | 0.309 | 0.318 | 0.336 |
| 70 | 0.372 | 0.392 | 0.435 | 0.326 | 0.337 | 0.361 | 0.313 | 0.323 | 0.342 | 0.308 | 0.316 | 0.334 | 0.305 | 0.313 | 0.329 |
| 80 | 0.364 | 0.383 | 0.422 | 0.320 | 0.331 | 0.354 | 0.309 | 0.318 | 0.336 | 0.304 | 0.312 | 0.328 | 0.301 | 0.309 | 0.324 |
| 90 | 0.357 | 0.375 | 0.412 | 0.316 | 0.326 | 0.348 | 0.305 | 0.314 | 0.331 | 0.300 | 0.308 | 0.324 | 0.298 | 0.305 | 0.320 |

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
International Conference, Chania, Crete, Greece, 8 - 11 June 2010

| | | | | | | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 100 | 0.352 | 0.368 | 0.403 | 0.313 | 0.322 | 0.342 | 0.302 | 0.310 | 0.327 | 0.298 | 0.305 | 0.320 | 0.295 | 0.302 | 0.316 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|

Table 4. Upper percentage points for Cochran's test statistic distribution in case of five samples with equal size n

| n | De(1) | | | De(2) | | | De(3) | | | De(4) | | | De(5) | | |
|-----|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|
| | α | | | α | | | α | | | α | | | α | | |
| | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 | 0.1 | 0.05 | 0.01 |
| 5 | 0.623 | 0.684 | 0.787 | 0.504 | 0.551 | 0.642 | 0.464 | 0.505 | 0.588 | 0.446 | 0.484 | 0.562 | 0.436 | 0.472 | 0.548 |
| 8 | 0.537 | 0.591 | 0.690 | 0.426 | 0.461 | 0.533 | 0.392 | 0.421 | 0.482 | 0.376 | 0.403 | 0.458 | 0.367 | 0.393 | 0.446 |
| 10 | 0.501 | 0.550 | 0.645 | 0.397 | 0.428 | 0.491 | 0.366 | 0.392 | 0.444 | 0.352 | 0.375 | 0.422 | 0.344 | 0.366 | 0.411 |
| 15 | 0.445 | 0.485 | 0.567 | 0.355 | 0.379 | 0.429 | 0.330 | 0.349 | 0.390 | 0.318 | 0.336 | 0.372 | 0.312 | 0.329 | 0.363 |
| 20 | 0.412 | 0.447 | 0.520 | 0.332 | 0.352 | 0.394 | 0.310 | 0.326 | 0.360 | 0.300 | 0.315 | 0.345 | 0.295 | 0.308 | 0.337 |
| 25 | 0.388 | 0.420 | 0.485 | 0.316 | 0.334 | 0.371 | 0.297 | 0.311 | 0.341 | 0.288 | 0.301 | 0.328 | 0.283 | 0.295 | 0.320 |
| 30 | 0.370 | 0.399 | 0.459 | 0.305 | 0.321 | 0.354 | 0.287 | 0.300 | 0.327 | 0.279 | 0.291 | 0.315 | 0.275 | 0.286 | 0.308 |
| 40 | 0.347 | 0.371 | 0.371 | 0.290 | 0.303 | 0.331 | 0.275 | 0.285 | 0.308 | 0.268 | 0.278 | 0.298 | 0.264 | 0.273 | 0.292 |
| 50 | 0.330 | 0.352 | 0.397 | 0.280 | 0.291 | 0.316 | 0.266 | 0.276 | 0.295 | 0.260 | 0.269 | 0.286 | 0.257 | 0.265 | 0.281 |
| 60 | 0.318 | 0.337 | 0.378 | 0.272 | 0.283 | 0.304 | 0.220 | 0.227 | 0.242 | 0.254 | 0.262 | 0.278 | 0.252 | 0.259 | 0.274 |
| 70 | 0.309 | 0.326 | 0.363 | 0.266 | 0.276 | 0.296 | 0.255 | 0.263 | 0.279 | 0.250 | 0.257 | 0.272 | 0.247 | 0.254 | 0.268 |
| 80 | 0.301 | 0.318 | 0.352 | 0.262 | 0.271 | 0.289 | 0.251 | 0.259 | 0.274 | 0.247 | 0.253 | 0.267 | 0.244 | 0.250 | 0.263 |
| 90 | 0.295 | 0.310 | 0.342 | 0.258 | 0.266 | 0.284 | 0.248 | 0.255 | 0.269 | 0.244 | 0.250 | 0.263 | 0.242 | 0.247 | 0.259 |

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
International Conference, Chania, Crete, Greece, 8 - 11 June 2010

| | | | | | | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 100 | 0.290 | 0.304 | 0.334 | 0.255 | 0.263 | 0.279 | 0.246 | 0.252 | 0.265 | 0.242 | 0.247 | 0.259 | 0.239 | 0.245 | 0.256 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|

Personalization of Text Information Retrieval with Bayesian Networks and Evolutionary Algorithms

Piotr Lipinski

Institute of Computer Science, University of Wrocław, Wrocław, Poland
(e-mail: lipinski@ii.uni.wroc.pl)

Abstract. Information retrieval focuses on searching through large databases of unstructural information and finding documents relevant to queries sent by users. Queries are usually imprecise and often do not reflect exactly some hidden user's intentions. In some cases, users send the same query many times or want to continuously monitor databases (changing frequently - each time when new documents appear or some others disappear) with the same query, which enables to collect some user's feedback, analyze it to detect hidden user's intentions and tune further searching.

In the approach proposed, each user's query has assigned a probability distribution over the document space, modelled by a Bayesian Network, which describes documents relevant to supposed user's expectations. At the beginning, the probability distribution is defined by the original user's query. After getting user's feedback, the probability distribution is updated to better fit hidden user's intentions by solving an optimization problem using an evolutionary algorithm, based on Bayesian Optimization Algorithm, estimating a new probability distribution. Information retrieval continues to search through databases with the new probability distribution and all the process repeats until the user stops monitoring databases.

Results of some experiments performed on real-life data prove that such a system is able to personalize information retrieval, adapt to hidden user's intentions and increase accuracy of results.

Keywords: Fitting models for data, Graphical models and Bayesian networks, Classification and Documentation, Data and Text Mining, Genetic and Fuzzy Algorithms.

1 Introduction

Information retrieval (IR) focuses on searching through large databases of unstructural information and finding documents relevant to queries sent by users [1]. Queries are usually imprecise and often do not reflect exactly some hidden user's intentions. In some cases, users send the same query many times or want to continuously monitor databases (changing frequently - each time when new documents appear or some others disappear) with the same query, which enables to collect some user's feedback, analyze it to detect hidden user's intentions and tune further searching.

In the approach proposed, each user's query has assigned a probability distribution over the document space, modelled by a Bayesian Network (BN) [3], which describes documents relevant to supposed user's expectations. At the beginning, the probability distribution is defined by the original user's query. After getting user's feedback, the probability distribution is updated to better fit hidden user's intentions by solving an optimization problem using an evolutionary algorithm, based

on Bayesian Optimization Algorithm (BOA) [6], [4], estimating a new probability distribution. IR continues to search through databases with the new probability distribution and all the process repeats until the user stops monitoring databases.

This paper is structured in the following manner: Section 2 introduces classic text information retrieval. Section 3 discusses its personalization. Section 4 presents an evolutionary algorithm to construct proper bayesian networks. Section 5 reports some experiments on real-life data. Finally, Section 6 concludes the paper.

2 Classic Text Information Retrieval

IR focuses on searching through large databases of unstructural information and finding documents relevant to queries sent by users. Information databases store descriptions of documents, such as text messages, on-line articles, news headlines or blog entries, where each document is described by some meta-data, such as authors, languages, publishers, categories or keywords, as well as some content-based data.

Users execute queries on information databases and receive lists of best matching documents. Queries consist of constraints on meta-data and sets of significant words for the content of documents. IR filters information databases according to the constraint on meta-data and searches through resulting documents for contents relevant to the set of significant words.

Information databases usually change with time, when new documents appear and some others disappear, so executing the same query in different time may lead to different results.

2.1 Information Database

Let \mathcal{D} denote the information database consisting of a number N of documents D_1, D_2, \dots, D_N . Each document $D_j, j = 1, 2, \dots, N$, is described by two vectors:

- $\mathbf{c}_j \in C_1 \times C_2 \times \dots \times C_K$, related to meta-data, where C_1, C_2, \dots, C_K denote domains of such meta-data components, and
- $\mathbf{d}_j \in \mathbb{R}^L$, related to content-based data, whose components correspond to significance of some preselected terms in the document, according to *the TF-IDF representation* [1], described further.

Let $\Omega_C = C_1 \times C_2 \times \dots \times C_K$ denote the meta-data space, $\Omega_D = \mathbb{R}^L$ denote the content-based data space, and $\Omega = \Omega_C \times \Omega_D$ denote the document space.

Let \mathcal{T} denote the list of some preselected terms consisting of a number L of terms t_1, t_2, \dots, t_L which includes all the significant words appearing in the entire information database \mathcal{D} (after simple preprocessing, such as lexical analysis, stopword elimination and stemming). For each term $t_i, i = 1, 2, \dots, L$, and each document $D_j, j = 1, 2, \dots, N$, let f_{ij} denote the number of occurrences of the term t_i in the document D_j , called *the term frequency* (TF) [1].

Term frequencies enable to compute *the term-document matrix* $\mathbf{D} \in \mathbb{R}^{L \times N}$ which describes relations between documents and terms [1]. For $i = 1, 2, \dots, L$, $j = 1, 2, \dots, N$, each element d_{ij} of the matrix \mathbf{D} is

$$d_{ij} = l_{ij}g_i, \quad (1)$$

where l_{ij} is a local factor that measure the importance of the term t_i in the document D_j and g_i is a global factor that measures the importance of the term t_i in the entire information database \mathcal{D} . In the TF-IDF representation, the local factor l_{ij} is the term frequency (TF), $l_{ij} = f_{ij}$, the global factor g_i is *the inverse document frequency* (IDF)

$$g_i = \log_2 \frac{N}{\sum_{j=1}^N \text{bin}(f_{ij})}, \quad (2)$$

where $\text{bin}(f_{ij}) = 1$ for $f_{ij} \neq 0$ and $\text{bin}(f_{ij}) = 0$ otherwise. Successive columns of the term-document matrix \mathbf{D} define the vectors $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N$ being the content-based description of the documents D_1, D_2, \dots, D_N .

2.2 Basic User's Query and Information Retrieval

In the classic approach, each query Q consists of two parts: a constraint on meta-data, $C \subset \Omega_C$, and a set of significant terms for the content of documents, $\mathcal{S} = \{s_1, s_2, \dots, s_l\} \subset \mathcal{T}$. Such a set \mathcal{S} defines an element $\mathbf{s} = (\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_L) \in \Omega_D$ in such a way that $\tilde{s}_i = 1$ if $t_i \in \mathcal{S}$ and $\tilde{s}_i = 0$ otherwise, for $i = 1, 2, \dots, L$.

IR tries to find all the documents D from the information database \mathcal{D} , which hold the constraint C and have the content similar to \mathbf{s} , i.e. all the documents $D = (\mathbf{c}, \mathbf{d}) \in \Omega$ such that

$$\mathbf{c} \in C \subset \Omega_C \quad \text{and} \quad \varrho(\mathbf{d}, \mathbf{s}) < \varepsilon, \quad (3)$$

where ϱ is a document similarity measure on the content-based data space, described further, and $\varepsilon > 0$ is a specific constant.

One of more popular document similarity measures is *the cosine measure* [1]. For $\mathbf{d}_1 \in \Omega_D$ and $\mathbf{d}_2 \in \Omega_D$ being the content-based description of two documents D_1 and D_2 , respectively, the cosine measure $\varrho(\mathbf{d}_1, \mathbf{d}_2)$ is the cosine of the angle between vectors \mathbf{d}_1 and \mathbf{d}_2 in the space \mathbb{R}^L ,

$$\varrho(\mathbf{d}_1, \mathbf{d}_2) = \frac{\mathbf{d}_1^T \mathbf{d}_2}{|\mathbf{d}_1| \cdot |\mathbf{d}_2|}, \quad (4)$$

where $|\mathbf{d}_1|, |\mathbf{d}_2|$ denote lengths of vectors $\mathbf{d}_1, \mathbf{d}_2$, respectively.

Figure 1 illustrates the document space Ω with documents from the information database \mathcal{D} , marked by dots, and a query $Q = (C, \mathbf{s})$, marked by a cross, as well as the result set – a set of documents holding the constraint C and having the content similar to \mathbf{s} .

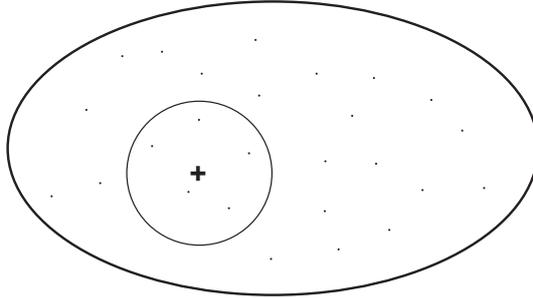


Fig. 1. The document space Ω with documents from the information database \mathcal{D} (marked by dots) and a query Q (marked by a cross) as well as the result set.

3 Personalized Text Information Retrieval

In some cases, users send the same query many times or want to continuously monitor information databases (changing frequently - each time when new documents appear and some others disappear) with the same query, which enables to improve information retrieval and provide more accurate results by collecting some user's feedback, analyzing it to detect hidden user's intentions and tuning further searching.

3.1 User's Feedback

Let \mathcal{R} denote the set of documents from the information database \mathcal{D} returned to the user after the initial query Q consisting of a number m of documents R_1, R_2, \dots, R_m . Each document $R_j, j = 1, 2, \dots, m$, is evaluated by the user by assigning a relevance factor $r_j \in \{0, 1\}$, where $r_j = 1$ stands for relevant documents and $r_j = 0$ for irrelevant ones (in practice, such relevance factors r_j are assigned automatically when the user open, or does not open, the document R_j).

3.2 Personalized User's Query and Information Retrieval

Such relevance factors enable to precise the initial user's query Q taking into consideration hidden user's intentions by assigning to the query Q a probability distribution P_Q over Ω describing documents relevant to supposed user's expectations.

P_Q is a $(K + L)$ -dimensional joint probability distribution and is modelled by a Bayesian Network (BN) [3] [5] [2], which is a directed acyclic graph, where nodes represents random variables and edges represents conditional dependencies between them (if two nodes are not connected by an edge, they are conditionally independent of each other). Each node has a marginal conditional probability distribution (in discreet cases - a table of values, in continuous cases - a density function), which describes distribution of the random variable for a particular set of values of parent nodes.

In order to avoid an excessively large BN (the number L of preselected terms usually exceeds 1000), in the approach proposed, P_Q is modelled by a BN with a directed acyclic graph $\mathcal{B} = (V, E)$, where the set of vertices $V = \{v_1, v_2, \dots, v_{K+l}\}$ consists of $K+l$ vertices corresponding to K meta-data components C_1, C_2, \dots, C_K and l terms s_1, s_2, \dots, s_l from the initial query Q and the set of edges $E \subset V \times V$ is initially empty (the edges as well as the marginal probability distributions are estimated further on the basis of user's feedback). In such an approach, the joint probability distribution P_Q modelled by the BN is

$$P_Q(D) = \prod_{j=1}^K P_Q(C_j | \pi_{C_j}) \prod_{j=1}^L P_Q(D_j | \pi_{D_j}), \quad (5)$$

where $P_Q(D_j | \pi_{D_j})$ has a uniform distribution for $t_j \notin S = \{s_1, s_2, \dots, s_l\}$ and π_V denotes parent vertices of the vertex V .

Figure 2 presents an example of BN for a query Q with a set of terms $S = \{s_1, s_2, s_3, s_4, s_5\}$, where $L = 5$ and $\Omega_D = \mathbb{R}^5$ as well as, for the sake of simplicity, $K = 0$ and $\Omega_C = \emptyset$. Figure 2 (A) presents an initial BN with no edges (variables are conditionally independent), and Figure 2 (B) presents a further BN with marginal conditional probability distributions in successive nodes S_1, S_2, \dots, S_5 corresponding to $P(S_1), P(S_2|S_1), P(S_3|S_1), P(S_4|S_1, S_3), P(S_5|S_1)$, respectively. In such an approach, the joint probability distribution P_Q modelled by the BN is

$$P_Q(D) = P(S_1) \cdot P(S_2|S_1) \cdot P(S_3|S_1) \cdot P(S_4|S_1, S_3) \cdot P(S_5|S_1). \quad (6)$$

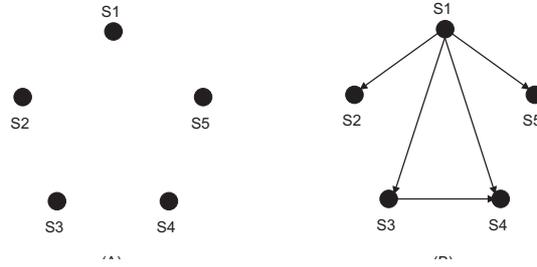


Fig. 2. A bayesian network for a query Q with a set of terms $S = \{s_1, s_2, s_3, s_4, s_5\}$ (for the sake of simplicity, $K = 0$ and $\Omega_C = \emptyset$): (A) all variables are conditionally independent, (B) S_2, S_3, S_5 depends on S_1 and S_4 depends on S_1 and S_3 .

4 Constructing Bayesian Networks

In order to construct the BN which models hidden user's intentions defined by user's feedback, an evolutionary algorithm, based on Bayesian Optimization Algorithm

(BOA) [6], [4], is applied. It deals with a bayesian network \mathcal{B} and a population $\mathcal{P} \subset \Omega$ of documents generated randomly using \mathcal{B} , as described in the previous section, with the aim to generate a set of documents relevant to hidden user's intentions. In order to simplify the approach, all the continuous domains of components of Ω were quantized.

For a document $D \in \mathcal{R}$, its relevance $F(D)$ to hidden user's intentions is defined by user's feedback. For a document $D \in \Omega \setminus \mathcal{R}$, its relevance $F(D)$ to hidden user's intentions is based on its distances to documents evaluated by the user in such a way that

$$F(D) = \max_{j=1,2,\dots,m} \frac{F(R_j)}{\varrho(D, R_j) + 1}. \quad (7)$$

Algorithm 1 presents the overview of the evolutionary algorithm, which maximizes the objective function F and builds a proper BN. It starts with taking an initial network \mathcal{B}_0 , generating a random initial population \mathcal{P}_0 of size n and evaluating it. If it is the first attempt to personalize the user's query, \mathcal{B}_0 is an empty network with no edges, otherwise, it is the network built in the previous attempt (with previous user's feedback).

Afterwards, the evolution starts with improving the population \mathcal{P}_t by removing a half of the weakest individuals and replacing them by a half of the strongest individuals, updating the network \mathcal{B}_t to model the current population \mathcal{P}_t , generating a random new population and evaluating it. Finally, the evolution process repeats until a termination condition is hold (normally, after a certain number of iterations).

Updating the network \mathcal{B}_t consists in adapting it to model the current population \mathcal{P}_t by a sequence of trials of adding, removing or changing a direction of a randomly chosen edge with the aim to maximize *the K2 metrics*. The K2 metrics is given by

$$K_2(\mathcal{B}|\mathcal{P}) = \prod_{i=1}^{K+l} \prod_{\pi_{V_i}} \frac{1}{\Gamma(m(\pi_{V_i}) + 1)} \cdot \prod_{v_i} \Gamma(m(v_i, \pi_{V_i}) + 1) \quad (8)$$

where the product over π_{V_i} runs over all instances of the parents of V_i , the product over v_i runs over all instances of V_i , $m(\pi_{V_i})$ denotes the number of instances in \mathcal{P} with the parents of V_i instantiated to π_{V_i} , and $m(\pi_{V_i})$ denotes the number of instances in \mathcal{P} with the parents of V_i instantiated to π_{V_i} and V_i instantiated to v_i .

5 Experiments

A number of experiments on real-life data were performed to validate the approach proposed. The information database consisted of descriptions of documents retrieved from popular news services (about 20 data sources, about 10000 documents in total, collected over a period of one year). Meta-data consisted of 10 components, leading to a preliminary document categorization. Content-based data consisted of 3000 components, corresponding to 3000 preselected terms.

Algorithm 1 Evolutionary Algorithm to Personalization of IR (EAPIR)

```

 $\mathcal{B}_0 = \text{Initial-Bayesian-Network}();$ 
 $\mathcal{P}_0 = \text{Random-Population}(\mathcal{B}_0, n);$ 
Population-Evaluation( $\mathcal{P}_0$ );
 $t = 0;$ 
while not Termination-Condition( $\mathcal{P}_t$ ) do
  Replace-Weakest-Individuals( $\mathcal{P}_t$ );
   $\mathcal{B}_{t+1} = \text{Update-Bayesian-Network}(\mathcal{B}_t, \mathcal{P}_t);$ 
   $\mathcal{P}_{t+1} = \text{Random-Population}(\mathcal{B}_{t+1}, n);$ 
  Population-Evaluation( $\mathcal{P}_{t+1}$ );
   $t = t + 1;$ 
end while
return  $\mathcal{P}_t;$ 

```

In the first part of experiments, each experiment concerned a user’s query and a manually selected set of documents, perfectly suiting hidden user’s intentions. The user’s query were processed by the system and a set of documents was returned to the user. The user evaluated it and user’s feedback returned to the system, which adapted the searching process and returned a new set of documents. Such a process was repeated until the user did not cancel the query.

Table 1 presents summary of results of these experiments for 3 iterations. Due to the long time necessary for manual preparing such data sets, only 10 experiments were performed. However, it is easy to see that personalization of IR enables to reduce the total number of documents found and returned to the user without significant loss in expected documents.

Table 1. Summary of results of experiments with manually selected set of documents perfectly suiting hidden user’s intentions.

| experiment | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| number of terms in the query | 5 | 5 | 10 | 10 | 10 | 10 | 20 | 20 | 20 | 20 |
| total number of documents (1st iter.) | 500 | 238 | 500 | 500 | 302 | 194 | 500 | 500 | 287 | 173 |
| number of expected documents (1st iter.) | 240 | 64 | 326 | 173 | 134 | 57 | 288 | 211 | 158 | 37 |
| total number of documents (2nd iter.) | 387 | 174 | 421 | 398 | 199 | 101 | 455 | 378 | 198 | 126 |
| number of expected documents (2nd iter.) | 238 | 64 | 311 | 162 | 138 | 57 | 267 | 209 | 149 | 35 |
| total number of documents (3rd iter.) | 297 | 132 | 394 | 231 | 174 | 86 | 328 | 319 | 173 | 73 |
| number of expected documents (3rd iter.) | 237 | 63 | 309 | 157 | 135 | 51 | 267 | 208 | 149 | 35 |

In the second part of experiments, 1000 queries were processed and the average user’s feedback was studied in a few successive iterations of the IR process. Table 2 presents a summary of results. Clearly, the personalization of IR enables to improve the average document evaluation.

Table 2. Average user's feedback in successive iterations of the IR process.

| number of terms in the query | 5 | 10 | 20 |
|---|--------|--------|--------|
| average document evaluation (1st iter.) | 0.3892 | 0.4726 | 0.5972 |
| average document evaluation (2nd iter.) | 0.5195 | 0.6294 | 0.7694 |
| average document evaluation (3rd iter.) | 0.7214 | 0.8016 | 0.8731 |

6 Conclusions

This paper addresses the problem of optimization of information retrieval by personalization of searching through the information database according to the hidden user's intentions. Experiments performed on real-life data proved that the approach presented were able to improve IR, adapt to hidden user's intentions and increase accuracy of results. EAs applied in the approach, based on the Bayesian Optimization Algorithm (BOA), were able to find efficient solutions for the optimization problem and construct a BN precisely describing user's feedback.

Such an improvement may be applied in the IR systems, where users send the same query many times or want to continuously monitor databases with the same query, such as corporate information databases with many users waiting for documents related to their domains and duties (accountants, analytics, investors or managers).

Further research may improve the approach proposed by additional studies on modelling the probability distribution and constructing the BN (for instance, adding additional nodes to the BN in the EA), experiments with dimensionality reduction in the IR as well as further validation on larger data sets.

References

1. Ghosh, J., Strehl, A., Similarity-Based Text Clustering: A Comparative Study, [in] Grouping Multidimensional Data, eds. J. Kogan, C. Nicholas, M. Teboulle, Springer, 2006, pp.73–97.
2. Heckerman, D., Geiger, D., Chickering, D. M., Learning Bayesian networks: The combination of knowledge and statistical data, Technical Report MSR-TR-94-09, Microsoft Research, Redmont, 1994.
3. Howard, R., Matheson, J., Influence diagrams, [in] The Principles and Applications of Decision Analysis, Volume II, eds. R. Howard, J. Matheson, Strategic Decisions Group, Menlo Park, 1984, pp. 719-762.
4. Larranaga, P., Lozano, J., Estimation of Distribution Algorithms, A New Tool for Evolutionary Computation, Kluwer Academic Publishers, 2002.
5. Pearl, J., Probabilistic reasoning in intelligent systems, Networks of plausible inference, Morgan Kaufmann, 1988.
6. Pelikan, M., Goldberg, D., Cantu-Paz, E., BOA: The Bayesian Optimization Algorithm, [in] Proceedings of Genetic and Evolutionary Computation Conference, 1999, pp.525-532.

Low-Rank Data Format for Uncertainty Quantification

Alexander Litvinenko and Hermann G. Matthies

Institute of Scientific Computing
Technische Universität Braunschweig, Braunschweig, Germany,
(e-mail: wire@tu-bs.de)

Abstract. We research how uncertainties in the input data (parameters, coefficients, right-hand sides, boundary conditions, computational geometry) spread of in the solution. Since all realisations of random fields are too much information, we demonstrate an algorithm of their low-rank approximation. This algorithm is based on singular value decomposition and has linear complexity. This low-rank approximation allows us to compute main statistics such as the mean value, variance, exceedance probability with a linear complexity and with drastically reduced memory requirements.

Keywords: uncertainty quantification, stochastic elliptic partial differential equations, Karhunen-Loève expansion, QR-algorithm, sparse data format, low-rank data format.

1 Introduction

Nowadays the trend of numerical mathematics is often trying to resolve inexact mathematical models by very exact deterministic numerical methods. The reason of this inexactness is that almost each mathematical model of a real world situation contains uncertainties in the coefficients, right-hand side, boundary conditions, initial data as well as in the computational geometry. All these uncertainties can affect the solution dramatically, which is, in its turn, also uncertain. The information of the interest usually is not the whole set of the solutions (too much data), but some other stochastic information: cumulative distribution function, density function, mean value, variance, exceedance probability etc.

We consider mathematical models described by stochastic partial differential equations (SPDEs), where uncertainties are represented as random fields. Efficient numerical solution of such SPDEs requires an appropriate discretisation of the deterministic operator as well as the stochastic fields. The total number degrees of freedom (dofs) of the discrete model of the SPDE is the product of dofs of the deterministic and stochastic discretisations and can be, even after application of the truncated Karhunen-Loève expansion (KLE) [7] and polynomial chaos expansion (PCE) of Wiener [11], very high. Therefore data sparse techniques for representation of input and output data (solution) are necessary for efficient representation and computation.

In this work we compress the set of output random fields via the algorithm based on the singular value decomposition. The short idea is as follows. Let Z be a number of stochastic realisations of the solution (e.g., number of Monte Carlo simulations or a number of collocation points). Let $\mathbf{v}_i \in \mathbb{R}^n$, $i = 1..Z$, stochastic realisations of the solution (without the mean value). We build from all vectors \mathbf{v}_i the matrix $W := [\mathbf{v}_1, \dots, \mathbf{v}_Z] \in \mathbb{R}^{n \times Z}$ and compute its low-rank approximation $\tilde{W} = AB^T$. For every new vector \mathbf{v}_{Z+1} an update for the matrices A and B is computed on the fly with a linear complexity. In the conclusion, we demonstrate examples from aerodynamic (influence of uncertainties in the angle of attack α , in the Mach number Ma and in the airfoil geometry on the solution - drag, lift, pressure and friction coefficients).

2 Discretisation techniques

By definition, the Karhunen-Loève expansion (KLE) of a random field $\kappa(x, \omega)$ is the following series [7]

$$\kappa(x, \omega) = E_\kappa(x) + \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell} \phi_\ell(x) \xi_\ell(\omega), \quad (1)$$

where $\xi_\ell(\omega)$ are uncorrelated random variables and $E_\kappa(x)$ is the mean value of $\kappa(x, \omega)$, λ_ℓ and ϕ_ℓ are the eigenvalues and the eigenvectors of problem

$$T\phi_\ell = \lambda_\ell \phi_\ell, \quad \phi_\ell \in L^2(\mathcal{G}), \ell \in \mathbb{N}, \quad (2)$$

and operator T is defined like follows

$$T : L^2(\mathcal{G}) \rightarrow L^2(\mathcal{G}), \quad (T\phi)(x) := \int_{\mathcal{G}} \text{cov}_\kappa(x, y) \phi(y) dy,$$

where $\text{cov}_\kappa(x, y)$ a given covariance function. Throwing away all unimportant terms in KLE, one obtains the truncated KLE, which is a sparse representation of the random field $\kappa(x, \omega)$. Each random variable ξ_ℓ can be approximated in a set of new independent Gaussian random variables (polynomial chaos expansions (PCE) of Wiener [3,11]), e.g.

$$\xi_\ell(\omega) = \sum_{\beta \in \mathcal{J}} \xi_\ell^{(\beta)} H_\beta(\boldsymbol{\theta}(\omega)), \quad (3)$$

where $\boldsymbol{\theta}(\omega) = (\theta_1(\omega), \theta_2(\omega), \dots)$, $\xi_\ell^{(\beta)}$ are coefficients, H_β , $\beta \in \mathcal{J}$, is a Hermitian basis and $\mathcal{J} := \{\beta | \beta = (\beta_1, \dots, \beta_j, \dots), \beta_j \in \mathbb{N}_0\}$ a multi-index set [8]. Computing the truncated PCE for each random variable in KLE, one can make representation of the random field even more sparse. Polynomial expansion (3) is also called the **response surface**.

Since Hermite polynomials are orthogonal, the coefficients $\xi_\ell^{(\beta)}$ can be computed by projection

$$\xi_\ell^{(\beta)} = \frac{1}{\beta!} \int_{\Theta} H_\beta(\boldsymbol{\theta}) \xi_\ell(\boldsymbol{\theta}) \mathbb{P}(d\boldsymbol{\theta}).$$

This multidimensional integral over Θ can be computed approximately, for example, on a sparse Gauss-Hermite grid

$$\xi_\ell^{(\beta)} = \frac{1}{\beta!} \sum_{i=1}^n H_\beta(\boldsymbol{\theta}_i) \xi_\ell(\boldsymbol{\theta}_i) w_i, \tag{4}$$

where weights w_i and points $\boldsymbol{\theta}_i$ are defined from sparse Gauss-Hermite integration rule.

After a finite element discretisation (see [5] for more details) the discrete eigenvalue problem (2) looks like

$$MCM\boldsymbol{\phi}_\ell = \lambda_\ell^h M\boldsymbol{\phi}_\ell, \quad C_{ij} = \text{cov}_\kappa(x_i, y_j). \tag{5}$$

Here the mass matrix M is stored in a usual data sparse format and the dense matrix $C \in \mathbb{R}^{n \times n}$ (requires $\mathcal{O}(n^2)$ units of memory) is approximated in the sparse \mathcal{H} -matrix format [5] (requires only $\mathcal{O}(n \log n)$ units of memory) or in the Kronecker low-rank tensor format [4]. To compute m eigenvalues ($m \ll n$) and corresponding eigenvectors we apply the Lanczos eigenvalue solver [6,10].

3 Data compression

Stochastic random fields require a large amount of memory and computational resources. The aim is to find a low-rank format for all presented input and output random fields and for each new realisation to compute only corresponding low-rank update (see, e.g. [1]). It can be practical when, e.g. many thousands Monte Carlo simulations are computed and stored. This low-rank data format makes memory requirements smaller and the computational process faster.

Let $\mathbf{v}_i \in \mathbb{R}^n$ be the solution vector (the deterministic component is subtracted), where $i = 1..Z$ a number of stochastic realisations of the solution. Build from all these vectors the matrix $W = (\mathbf{v}_1, \dots, \mathbf{v}_Z) \in \mathbb{R}^{n \times Z}$. Consider the factorization

$$W = AB^T \quad \text{where } A \in \mathbb{R}^{n \times k} \quad \text{and} \quad B \in \mathbb{R}^{Z \times k}. \tag{6}$$

Definition 1. We say that matrix W is a **rank- k matrix** if the representation (6) is given. We denote the class of all rank- k matrices for which factors A and B^T in (6) exist by $\mathcal{R}(k, n, Z)$. If $W \in \mathcal{R}(k, n, Z)$ we say that W has a **low-rank representation**.

The first aim is to compute a rank- k approximation \tilde{W} of W , such that

$$\|W - \tilde{W}\| < \varepsilon, \quad k \ll \min\{n, Z\}.$$

The second aim is to compute an update for the approximation \tilde{W} with a linear complexity for every new coming vector \mathbf{v}_{Z+1} . Below we present the algorithm which does this.

To get the reduced singular value decomposition we omit all singular values, which are smaller than some level ε or, alternative variant, we leave a fixed number of largest singular values. After truncation we speak about *reduced singular value decomposition (denoted by rSVD)* $\tilde{W} = \tilde{U} \tilde{\Sigma} \tilde{V}^T$, where $\tilde{U} \in \mathbb{R}^{n \times k}$ contains the first k columns of U , $\tilde{V} \in \mathbb{R}^{Z \times k}$ contains the first k columns of V and $\tilde{\Sigma} \in \mathbb{R}^{k \times k}$ contains the k -biggest singular values of Σ . There is Lemma (see more in [9] or [2]) which tells that matrix \tilde{W} is the best approximation of W in the class of all rank- k matrices.

The computation of such basic statistics as the mean value, the variance, the exceedance probability can be done with a linear complexity. The following examples illustrate computation of the mean value and the variance.

Let $W = (\mathbf{v}_1, \dots, \mathbf{v}_Z) \in \mathbb{R}^{n \times Z}$ and its rank- k representation $W = AB^T$, $A \in \mathbb{R}^{n \times k}$, $B^T \in \mathbb{R}^{k \times Z}$ be given. Denote the j -th row of matrix A by $\mathbf{a}_j \in \mathbb{R}^k$ and the i -th column of matrix B^T by $\mathbf{b}_i \in \mathbb{R}^k$.

1. One can compute the mean solution $\bar{\mathbf{v}} \in \mathbb{R}^n$ as follows

$$\bar{\mathbf{v}} = \frac{1}{Z} \sum_{i=1}^Z \mathbf{v}_i = \frac{1}{Z} \sum_{i=1}^Z A \cdot \mathbf{b}_i = A \bar{\mathbf{b}}, \quad (7)$$

The computational complexity is $\mathcal{O}(k(Z+n))$, besides $\mathcal{O}(nZ)$ for usual dense data format.

2. One can compute the mean value of the solution in a grid point x_j as follows

$$\bar{\mathbf{v}}(x_j) = \frac{1}{Z} \sum_{i=1}^Z \mathbf{v}_i(x_j) = \frac{1}{Z} \sum_{i=1}^Z \mathbf{a}_j \cdot \mathbf{b}_i = \mathbf{a}_j \bar{\mathbf{b}}. \quad (8)$$

The computational complexity is $\mathcal{O}(kZ)$.

3. One can compute the variance of the solution $\text{var}(\mathbf{v}) \in \mathbb{R}^n$ by the computing the covariance matrix and taking its diagonal. First, we compute the centred matrix $W_c := W - \bar{W}e^T$, where $\bar{W} = W \cdot e/Z$ and $e = (1, \dots, 1)^T$. Computing W_c costs $\mathcal{O}(k^2(n+Z))$ (addition and truncation of rank- k matrices). By definition, the covariance matrix is $\text{cov} = W_c W_c^T$. The reduced singular value decomposition of W_c is $W_c = U \Sigma V^T$, $\Sigma \in \mathbb{R}^{k \times k}$, can be computed with a linear complexity via the QR algorithm (Section 3.1). Now, the covariance matrix can be written like

$$\text{cov} = W_c W_c^T = U \Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T. \quad (9)$$

The variance of the solution vector (i.e. the diagonal of the covariance matrix in (9)) can be computed with the complexity $\mathcal{O}(k^2(Z+n))$.

4. One can compute the variance value $\text{var}(\mathbf{v}(x_j))$ in a grid point x_j with a linear computational cost.
5. To compute minimum or maximum of the solution in a point x_j over all realisations cost $\mathcal{O}(kZ)$.

3.1 Low-rank update with linear complexity

Let $W = AB^T \in \mathbb{R}^{n \times Z}$ and matrices A and B be given.

An rSVD $W = U\Sigma V^T$ can be computed efficiently in three steps (QR algorithm for computing the reduced SVD):

1. Compute (reduced) a QR -factorization of $A = Q_A R_A$ and $B = Q_B R_B$, where $Q_A \in \mathbb{R}^{n \times k}$, $Q_B \in \mathbb{R}^{Z \times k}$, and upper triangular matrices $R_A, R_B \in \mathbb{R}^{k \times k}$.
2. Compute an reduced SVD of $R_A R_B^T = U' \Sigma V'^T$.
3. Compute $U := Q_A U'$, $V := Q_B V'^T$.

The first and third steps need $\mathcal{O}((n+Z)k^2)$ operations and the second step needs $\mathcal{O}(k^3)$. The total complexity of rSVD is $\mathcal{O}((n+Z)k^2 + k^3)$.

Suppose we have already matrix $W = AB^T \in \mathbb{R}^{n \times Z}$ containing solution vectors. Suppose also that matrix $W' \in \mathbb{R}^{n \times m}$ contains new m solution vectors. For the small matrix W' , computing the factors C and D^T such that $W' = CD^T$ is not expensive. Now our purpose is to compute with a linear complexity the new matrix $W_{\text{new}} \in \mathbb{R}^{n \times (Z+m)}$ in the rank- k format. To do this, we build two concatenated matrices $A_{\text{new}} := [A \ C] \in \mathbb{R}^{n \times 2k}$ and $B_{\text{new}}^T = \text{blockdiag}[B^T \ D^T] \in \mathbb{R}^{2k \times (Z+m)}$. Note that the difficulty now is that matrices A_{new} and B_{new} have rank $2k$. To truncate the rank from $2k$ to k we use the QR-algorithm above. Obtain

$$W_{\text{new}} = U \Sigma V^T = U (V \Sigma^T)^T = A_{\text{new}} B_{\text{new}}^T,$$

where $A_{\text{new}} \in \mathbb{R}^{n \times k}$ and $B_{\text{new}}^T \in \mathbb{R}^{k \times (Z+m)}$. Thus, the “update” of the matrix W is done with a linear complexity $\mathcal{O}((n+Z)k^2 + k^3 + (n+Z)k^2)$.

4 Numerics

Further numerical results are obtained in the MUNA project. We demonstrate the influence of uncertainties in the angle of attack, the Mach number and the airfoil geometry on the solution (the lift, drag, lift coefficient and skin friction coefficient). As an example we consider two-dimensional RAE-2822 airfoil. The deterministic solver is the TAU code with k-w turbulence model. We assume that α and Ma are gaussian with means $\bar{\alpha} = 2.79$, $\bar{Ma} = 0.734$ and the standard deviations $\sigma(\alpha) = 0.1$ and $\sigma(Ma) = 0.005$.

Table 1 demonstrates application of the collocation method computed in grid points of the sparse Gauss-Hermite two-dimensional grid ($Z = 5$ deterministic evaluations). The Hermite polynomials are of order 1 with two random variables (see (3)). In the last column we compute the measure of uncertainty σ/mean . It shows that 3.6% and 0.7% of uncertainties in α and in Ma correspondingly result in 2.1% and 15.1% of uncertainties in the lift CL and drag CD .

In Fig.1 we compare the cumulative distribution and density functions for the lift and drag, obtained via the response surface (PCE of order 1) and via 6360 Monte Carlo simulations. To get the large sample we evaluated 10^6 MC points on the response surface. Thus, one can see that very cheap collocation method (13 or 29 deterministic evaluations) produces similar to MC method with 6360 simulations. But, at the same time we can not say which result is more precise. The exact solution is unknown and 6360 MC simulations are too few.

| | mean | st. dev. σ | σ/mean |
|----------|--------|-------------------|----------------------|
| α | 2.79 | 0.1 | 0.036 |
| Ma | 0.734 | 0.005 | 0.007 |
| CL | 0.853 | 0.018 | 0.021 |
| CD | 0.0206 | 0.0031 | 0.151 |

Table 1. Uncertainties in the input parameters (α and Ma) and in the solution (CL and CD). PCE of order 1 and sparse Gauss-Hermite grid with 5 points.

The graphics in Fig. 2 demonstrate error bars $[\text{mean} - \sigma, \text{mean} + \sigma]$, σ the standard deviation, for the pressure coefficient cp and absolute skin friction cf in each surface point of the RAE2822 airfoil. The data are obtained from 645 realisation of the solution. One can see that the largest error occur at the shock ($x \approx 0.6$). A possible explanation is that the shock position is expected to slightly change with varying parameters α and Ma .

To decrease numerical complexity we build from all $Z = 645$ realisations of the solution the matrix W :

$$W = [\text{density}; \text{pressure}; \text{cp}; \text{cf}]^T \in \mathbb{R}^{2048 \times 645} \quad (10)$$

and compute its rank- k approximation \tilde{W} . The solution vector $\mathbf{v} \in \mathbb{R}^{2048}$ consists of four vectors: density, pressure, cp and cf . In Table 2 one can see fast decay of the approximation error. Additionally, one can also see very smaller memory requirement (dense matrix format costs 10.6MB).

We model uncertainties in the geometry of RAE2822 airfoil via random boundary perturbations:

$$\partial\mathcal{G}_\varepsilon(\omega) = \{x + \varepsilon\kappa(x, \omega)n(x) : x \in \partial\mathcal{G}\}, \quad (11)$$

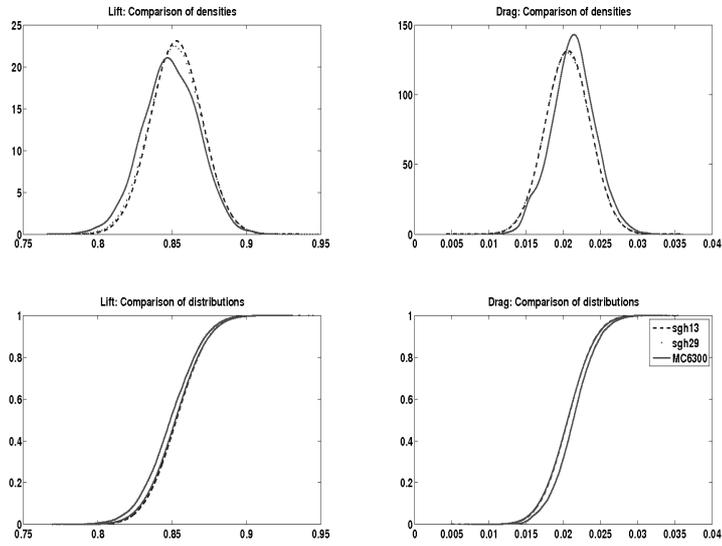


Fig. 1. Density functions (first row), cumulative distribution functions (second row) of CL (left) and CD (right). PCE is of order 1 with two random variables. Three graphics computed with 6360 MC simulations, 13 and 29 collocation points.

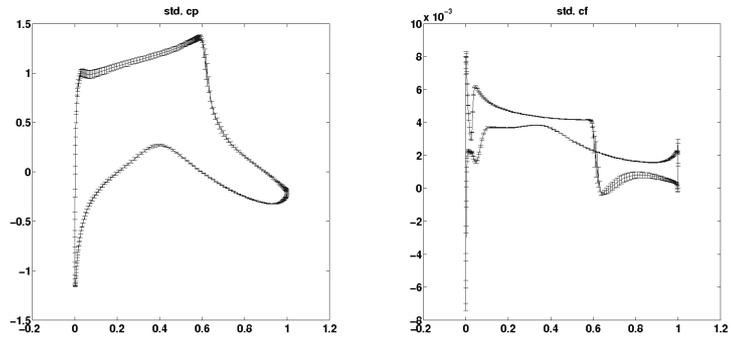


Fig. 2. Error bars [mean $- \sigma$, mean $+ \sigma$], σ standard deviation, in each point of RAE2822 airfoil for the cp and cf.

| rank k | 1 | 2 | 5 | 10 | 20 | 50 |
|---------------------------|------|------|-----|------|------|--------|
| $\ W - W_k\ _2 / \ W\ _2$ | 0.82 | 0.21 | 0.4 | 5e-3 | 5e-4 | 1.2e-5 |
| memory, kB | 22 | 43 | 108 | 215 | 431 | 1080 |

Table 2. Accuracy and memory requirements of the rank- k approximation of the solution matrix $W = [\text{density}; \text{pressure}; \text{cp}; \text{cf}]^T \in \mathbb{R}^{2048 \times 645}$.

where $n(x)$ is the normal vector in point x , $\kappa(x, \omega)$ a random field, \mathcal{G} the computational geometry and $\varepsilon \ll 1$. We assume that the covariance function is of Gaussian type

$$\text{cov}_\kappa(p_1, p_2) = \sigma^2 \cdot \exp(-d^2), \quad d = \sqrt{|x_1 - x_2|^2/l_1^2 + |z_1 - z_2|^2/l_2^2},$$

where $\sigma = 10^{-3}$, $p_1 = (x_1, 0, z_1)$, $p_2 = (x_2, 0, z_2)$, the covariance lengths $l_1 = |\max_i(x) - \min_i(x)|/10$ and $l_2 = |\max_i(z) - \min_i(z)|/10$. We took $m = 3$ KLE terms (1), stochastic dimension is 3 and the number of sparse Gauss-Hermite points (in 3d) for computing PCE coefficients (4) is 25. After building the response surfaces for CL and CD , we use them to generate 10^6 MC realisations. We obtain surprisingly small uncertainties in the CL and CD — 0.58% and 0.65% correspondingly. A possible explanation can be a small uncertain perturbations in the airfoil geometry.

References

1. M. Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra Appl.*, 415(1):20–30, 2006.
2. G. H. Golub and Ch. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.
3. T. Hida, Hui-Hsiung Kuo, J. Potthoff, and L. Streit. *White noise - An infinite-dimensional calculus*, volume 253 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.
4. B. N. Khoromskij and A. Litvinenko. Data sparse computation of the Karhunen-Loève expansion. *Numerical Analysis and Applied Mathematics: International Conference on Numerical Analysis and Applied Mathematics, AIP Conf. Proc.*, 1048(1):311–314, 2008.
5. B. N. Khoromskij, A. Litvinenko, and H. G. Matthies. Application of hierarchical matrices for computing the Karhunen-Loève expansion. *Computing*, 84(1-2):49–67, 2009.
6. C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Research Nat. Bur. Standards*, 45:255–282, 1950.
7. M. Loève. *Probability theory I. Graduate Texts in Mathematics, Vol. 45, 46*. Springer-Verlag, New York, fourth edition, 1977.
8. H. G. Matthies. Uncertainty quantification with stochastic finite elements. 2007. Part 1. Fundamentals. *Encyclopedia of Computational Mechanics*.
9. L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford Ser. (2)*, 11:50–59, 1960.
10. Y. Saad. *Numerical methods for large eigenvalue problems*. Algorithms and Architectures for Advanced Scientific Computing. Manchester University Press, Manchester, 1992.
11. N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

The Homogeneous Markov System (HMS) as an elastic medium. The three-dimensional case.

MAAITA J.-O.^{*}, TSAKLIDIS G.[†] and MELETLIDOU E.[‡]
jmaay@physics.auth.gr, tsaklidi@math.auth.gr, efthymia@auth.gr

Abstract

Every attainable structure of the so called continuous-time Homogeneous Markov System (HMS) with fixed size and state space $S=\{1,2,\dots,n\}$ is considered as a particle of \mathbf{R}^n and consequently the motion of the structure corresponds to the motion of the particle. Under the assumption that "the motion of every particle-structure at every time point is due to its interaction with its surroundings", \mathbf{R}^n becomes a continuum [9]. Then the evolution of the set of the attainable structures corresponds to the motion of the continuum. For the case of a three-state HMS it is stated that the concept of the two-dimensional isotropic elasticity can further interpret its evolution.

Keywords: Continuous time Markov system; Stochastic (population) systems; Isotropic elastic continuum.

1 Introduction

There are many applications of Markov systems reported in the literature, in areas of manpower planning, statistical physics, chemistry, demography, geography as well as in economics and health care planning. In looking for example applications of Homogeneous or non-Homogeneous Markov systems (or semi-Markov systems) reference could be given to student enrolment in universities[1], occupational mobility [2] and sea pollution [3], among others, while basic results concerning continuous time Markov models in manpower systems can be found in [4]-[7]. Main problems of interest regarding Homogeneous Markov Systems (HMSs) are their asymptotic behaviour, stability, asymptotic stability, control, variability, estimation, attainability, maintainability and entropy.

Consider a continuous-time HMS with state space $S=\{1,2,\dots,n\}$. The members of the system could be particles, biological organisms, parts of human population, etc. Every member of the system may be in one and only one of the states 1, 2,..., n, at some time point t, and it can move from some state

^{*}Postal address: Depart. of Physics, Arist. Univ. of Thessal., 54124, Thessaloniki, Greece.

[†]Postal address: Depart. of Mathem., Arist. Univ. of Thessal., 54124, Thessal., Greece.

[‡]Postal address: Depart. of Physics, Arist. Univ. of Thessal., 54124, Thessaloniki, Greece.

i to some other state j in the time interval $[t, \Delta t]$ with transition probability $p_{ij}[\Delta t]$, for every $t \in \mathbf{R}^+$. Then, every attainable structure of the continuous time HMS with n states and fixed size is considered as a point-particle of \mathbf{R}^n . Thus, the motion of an attainable structure corresponds to the motion of the respective point-particle in \mathbf{R}^n . Under the assumption that the motion of every particle at every time point is due to its interaction with its surroundings, \mathbf{R}^n is further seen as a continuum [9]-[10] and, hence, the evolution of the set of the HMS attainable structures corresponds to the deformation of the continuum. This turns to be a realistic assumption, since the motion of every point-particle depends entirely on its position in \mathbf{R}^n .

Under these considerations, the concepts of the state of stress and the relevant stress tensor can be associated with an n-dimensional HMS and, as far as the present paper is concerned, these are initially detailed in an example application dealing with a three-dimensional HMS. Then, given the rate of transition probabilities matrix of the HMS, a question is raised on whether the set of the attainable structures of the continuous time HMS may be considered as an elastic solid and, in this context, it is further examined whether the deformation of such a model could explain the evolution of the HMS. The study follows the steps of the methodology presented in [10], where the search for an answer to this question gave rise to the concept of multidimensional, anisotropic linear elasticity. The evolution of a 2-dimensional HMS was successfully interpreted in [10] through the deformation of a linearly elastic rod, while it was further mentioned that the evolution of an n-dimensional HMS may be interpreted through the deformation of an (n-1)-dimensional, anisotropic, linearly elastic solid.

Apart from the above concepts, the present paper develops further the three-dimensional HMS example application, and the evolution of the HMS is interpreted through the deformation of some two-dimensional isotropic elastic solid. The increased number of dimensions, as compared with the number of dimensions considered in [10], results into an increase of the number of the PDEs describing the motion of the present HMS and, consequently, it complicates the associated calculations. Using the field equations of elasticity, an explicit form of the stress tensor involved can still be evaluated analytically. It is therefore concluded that, under certain assumptions, the evolution of a three-dimensional HMS may successfully be interpreted through the deformation of a two-dimensional elastic solid material.

The successful interpretation of the evolution of HMSs through the deformation laws of elastic solids gives rise to further fruitful thoughts regarding the manner in which well known concepts and features met in classical and finite elasticity (e.g., anisotropy, strain energy) may be associated to HMSs and be exploited appropriately.

2 The continuous time Homogeneous Markov System as an elastic medium

For a continuous time HMS it is assumed that the transition probability of moving from some state i to j in the time interval $(t, t+\Delta t)$ is given by the relation

$$p_{ij}(t, t + \Delta t) = q_{ij}\Delta t + o(\Delta t), \quad (1)$$

where $o(\Delta t)$ denotes a quantity that becomes negligible when compared to Δt as $\Delta t \rightarrow 0$, that is $\lim_{\Delta t \rightarrow 0} (o(\Delta t)/\Delta t) = 0$. In the general case of non-homogeneous Markov systems the transition intensities q_{ij} may be time dependent.

In what follows let $\mathbf{x}(t) = (x_i(t))_{i \in S}$ denote the $n \times 1$ state vector of the HMS, the i -th component of which is the probability of a systems' member to possess state i at time t . Then the probabilistic law for the transitions given in (1) leads to the equation

$$x_j(t + \delta t) = x_i(t)(\delta_{ij} + q_{ij}\Delta t) + o(\Delta t), \quad (2)$$

where repeated indices denote summation over their range, and δ_{ij} stands for the Kronecker delta, having the value 1 when $i = j$ and 0 otherwise. From (2) the Kolmogorov equation can be derived, i.e.

$$\dot{\mathbf{x}}^T(t) = \mathbf{x}^T(t) \cdot \mathbf{Q}, \quad (3)$$

where $\dot{\mathbf{x}}(t)$ denotes the derivative of the vector $\mathbf{x}(t)$ with respect to t , $\mathbf{Q} = (q_{ij})_{i, j \in S}$ is the matrix of the transition intensities and the superscript T denotes transposition of the respective vector (or matrix).

Equation (3) represents the motion of a stochastic structure in \mathbf{R}^n . If we consider every structure of the HMS moving according to (3) as a 'particle' of the n -dimensional space \mathbb{R}^n we can assign material behavior to \mathbb{R}^n . From (3) we conclude that the velocity $\dot{\mathbf{x}}(t)$ of each particle depends only on the position. So we can assume that the motion of every particle, at every time t , depends on the interaction of that particle with its surroundings. Thus the HMS may be considered as a continuum moving according to equation (3).

Now, from (3) we get the trajectory of every initial HMS's structure $\mathbf{x}(0)$ moving in \mathbf{R}^n is given by

$$\mathbf{x}^T(t) = \mathbf{x}^T(0)e^{\mathbf{Q}t}, \quad t \geq 0. \quad (4)$$

As the initial state vectors $\mathbf{x}(0)$ run over all stochastic n -tuples, we get the respective set of the solutions $\mathbf{x}(t)$ given by (4), which is denoted by A_t and called "the set of the attainable structures". Let $A_n(t)$ be the region of \mathbf{R}^n defined by A_t . We are interested in the motion-evolution of the continuum possessing the region $A_n(0) \subset \mathbf{R}^n$ at time $t = 0$ in the velocity field defined by (3).

Now, equation (3) represents a system of n linear differential equations (DEs). Because of the stochasticity condition

$$x_1(t) + x_2(t) + \dots + x_n(t) = 1$$

the variables $x_i(t)$, $i = 1, 2, \dots, n$, are dependent and the motion takes place on the hyperplane

$$(\Pi) : x_1 + x_2 + \dots + x_n = 1.$$

In order to express the motion taking place on the ((n-1)-dimensional) hyperplane (Π) using only n-1 coordinates, we introduce a new coordinate system as follows. Firstly we assume, without loss of generality, that \mathbf{Q} is an irreducible matrix. In this case a stochastic stability point, π , exists, for which $\pi^T \cdot \mathbf{Q} = \mathbf{0}^T$. Consider at π a new orthogonal coordinate system $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$ where $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{(n-1)}$ belong to the hyperplane (Π) and $\mathbf{f}_n \perp (\Pi)$, and let

$$\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n] = [\mathbf{F}_1 | \mathbf{f}_n],$$

where $\mathbf{F}_1 = [\mathbf{f}_1 | \mathbf{f}_2 | \dots | \mathbf{f}_{(n-1)}]$. Equation (3) expressed with respect to the coordinate system $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{(n-1)}\}$, with origin at π , becomes $\dot{\mathbf{z}}^T(t) = \mathbf{z}^T(t) \cdot \mathbf{G}$, or simply

$$\dot{\mathbf{z}}^T = \mathbf{z}^T \cdot \mathbf{G}, \quad (5)$$

where $\mathbf{z} = (z_1, z_2, \dots, z_{(n-1)})^T$ and

$$\mathbf{G} = \mathbf{F}_1^T \mathbf{Q} \mathbf{F}_1. \quad (6)$$

The system of the n DEs of (3) is now reduced to the equivalent system of the (n-1) DEs given in (5). So, equation (5) can be used instead of (3) in order to study the dynamical evolution-motion of the HMS-continuum taking place on (Π). Note that, since $tr \mathbf{G} = tr \mathbf{Q} < 0$, the field defined by (5) is compressible.

Now, every part of the "material continuum" $A_n(t)$, $t \geq 0$, is supposed to be subjected to surface forces. Then the $n \times 1$ stress vector $\mathbf{t}^n(t)$ is defined at every point P enclosed by the infinitesimal surface S, where \mathbf{n} is the $n \times 1$ outward unit normal of the surface element ΔS of S. The state of stress at P is given by the set $\{\mathbf{t}^n\}$ generated from all the unit vectors \mathbf{n} , according to the formula

$$\mathbf{t}^n = \mathbf{T} \cdot \mathbf{n}$$

where \mathbf{T} is the symmetric $n \times n$ stress tensor.

The stress tensor $\mathbf{T} = (T_{ij}(\mathbf{z}, t))$, $i, j = 1, 2, \dots, n - 1$, should satisfy Cauchy's equation of motion

$$\rho(\mathbf{z}, t) \cdot \mathbf{a}(\mathbf{z}, t) = div \mathbf{T}(\mathbf{z}, t) + \mathbf{b}(\mathbf{z}, t), \quad (7)$$

at every point P of the medium, where \mathbf{z} is the position vector of P (with respect to the new coordinate system), $\rho(\mathbf{z}, t)$ is the density on the neighborhood of P at time t, $\mathbf{a}(\mathbf{z}, t)$ is the acceleration at P at time t and $\mathbf{b}(\mathbf{z}, t)$ represents a vector of possible body forces given by the description of a particular HMS.

The acceleration $\mathbf{a}(\mathbf{z}, t)$ appearing in (7) is given by

$$\mathbf{a}(\mathbf{z}, t) = \frac{\partial \mathbf{v}}{\partial t} + \nabla \mathbf{v} \cdot \mathbf{v},$$

where the (i,j)-entry of the $(n-1) \times (n-1)$ matrix $\nabla \mathbf{v}$ equals $\frac{\partial v_i}{\partial z_j}$. Since, by (5), the velocity is time independent, we get

$$\mathbf{a}(\mathbf{z}, t) = \nabla \mathbf{v} \cdot \mathbf{v} = \mathbf{G}^T \cdot \dot{\mathbf{z}} = (\mathbf{G}^T)^2 \cdot \mathbf{z} .$$

Thus

$$\mathbf{a}(\mathbf{z}, t) = \mathbf{a}(\mathbf{z}) = (\mathbf{G}^2)^T \cdot \mathbf{z} . \quad (8)$$

Let $E = (\varepsilon_{ij})$ be the $(n-1) \times (n-1)$ Eulerian strain tensor with

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial z_j} + \frac{\partial u_j}{\partial z_i} - \frac{\partial u_i}{\partial z_j} \frac{\partial u_j}{\partial z_i} \right) , \quad (9)$$

where $\mathbf{u} = (u_i)$ represents the displacement vector. Since the features of the HMS give no rise to consider it as an inhomogeneous or anisotropic medium, we will focus attention on the case of a homogeneous isotropic elastic continuum. For this case the stress tensor becomes

$$T_{ij} = \lambda \varepsilon_{kk} \Delta_{ij} + 2\mu \varepsilon_{ij} , \quad (10)$$

where λ and μ are Lamé constants.

With the use of the Lamé constants we can define two other constants that characterize a continuum: The Young modulus, that is the ratio of stress over strain towards axes x_1

$$E = \frac{T_{11}}{E_{11}} = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}$$

and the Poisson coefficient

$$\nu = -\frac{E_{22}}{E_{11}} = -\frac{E_{33}}{E_{11}} = \frac{\lambda}{2(\lambda + \mu)} .$$

3 The case of the 3-D continuous-time HMS

For the case of the three-dimensional ($S=\{1,2,3\}$) irreducible HMS, the intensity matrix has the form

$$\mathbf{Q} = \begin{pmatrix} -q_{12} - q_{13} & q_{12} & q_{13} \\ q_{21} & -q_{21} - q_{23} & q_{23} \\ q_{31} & q_{32} & -q_{31} - q_{32} \end{pmatrix} , \quad (11)$$

where $q_{ij} \geq 0$ for $i \neq j$, and the diagonal elements are not equal to 0. The stability point π of the HMS is the stochastic, left eigenvector of the intensity matrix (11) associated with the eigenvalue 0, i.e. $\pi^T \cdot \mathbf{Q} = \mathbf{0}^T$ and $\pi^T \cdot \mathbf{1} = \mathbf{1}^T$, where $\mathbf{1}$ is the column vector of 1's.

The base vectors of the orthogonal coordinate system $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$, with origin at π , can be chosen to be

$$\mathbf{f}_1 = \left(-\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right), \mathbf{f}_2 = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \mathbf{f}_3 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right) . \quad (12)$$

Then

$$\mathbf{F}_1 = \begin{pmatrix} -\sqrt{\frac{2}{3}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}. \quad (13)$$

According to (5), the motion of a particle-structure on the 2-dimensional hyperplane (Π) is expressed by the equation

$$\mathbf{z}^T(t) = (z_1(t), z_2(t)) = \mathbf{z}^T(0)e^{\mathbf{G}t}. \quad (14)$$

Now, since

$$\mathbf{u}(\mathbf{z}; t, t + \Delta t) = \mathbf{z}(t + \Delta t) - \mathbf{z}(t), \quad (15)$$

the entries ε_{ij} of the strain tensor can be found using (9).

In order to examine if the 3-D HMS can be interpreted as a homogeneous elastic medium we have to check if Cauchy's equation of motion, (7), is justified while substituting for the required acceleration and density, using (8) and the continuity equation

$$\frac{\partial \rho}{\partial t} + \rho \cdot \text{div}(\mathbf{u}) = 0. \quad (16)$$

The mass forces appearing in (7) to meet the general case, are set equal to $\mathbf{0}$.

4 An illustrative example

Consider a closed continuous-time HMS with state space $S=\{1,2,3\}$ and intensity matrix

$$\mathbf{Q} = \begin{pmatrix} -4.7 & 4 & 0.7 \\ 4.02 & -4.22 & 0.2 \\ 0.2 & 2 & -2.2 \end{pmatrix}.$$

The stability point π of the HMS is the stochastic, left eigenvector of the intensity matrix (11) associated with the eigenvalue 0, that is $\pi^T \cdot \mathbf{Q} = \mathbf{0}^T$ and $\pi^T \cdot \mathbf{1} = \mathbf{1}^T$, where $\mathbf{1}$ is the column vector of 1's. It is found that $\pi^T = (0.389, 0.447, 0.164)$.

The base vectors of the orthogonal coordinate system $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$, with origin at π , can be chosen to be those given by (12). Then, by (13) and (6), the matrix \mathbf{G} appearing in the reduced matrix equation (5), which expresses the motion on the hyperplane (Π), is found to be

$$\mathbf{G} = \mathbf{F}_1^T \mathbf{Q} \mathbf{F}_1 = \begin{pmatrix} -6.81 & 1.96876 \\ 3.30822 & -4.31 \end{pmatrix},$$

with eigenvalues

$$\lambda_1 = -8.40716, \quad \lambda_2 = -2.71824.$$

Since $\lambda_1, \lambda_2 < 0$, the velocity field $\dot{\mathbf{z}}^T = \mathbf{z}^T \cdot \mathbf{G}$ is compressible.

Now, using (14), we get the equations of motion

$$\begin{aligned} z_1(t) &= (0.72e^{-8.402t} + 0.28e^{-2.718t}) z_{10} + (-0.582e^{-8.402t} + 0.582e^{-2.718t}) z_{20} \\ z_2(t) &= (-0.346e^{-8.402t} + 0.346e^{-2.718t}) z_{10} + (0.28e^{-8.402t} + 0.72e^{-2.718t}) z_{20}. \end{aligned}$$

Then from (15) we derive the components of the displacement vector:

$$\begin{aligned} u_1(\mathbf{z}; t, t + \Delta t) &= (-1 + 0.72e^{-8.402\Delta t} + 0.28e^{-2.718\Delta t}) z_1 \\ &\quad + (-0.582e^{-8.402\Delta t} + 0.582e^{-2.718\Delta t}) z_2 \\ u_2(\mathbf{z}; t, t + \Delta t) &= (-0.346e^{-8.402\Delta t} + 0.346e^{-2.718\Delta t}) z_1 \\ &\quad + (-1 + 0.28e^{-8.402\Delta t} + 0.72e^{-2.718\Delta t}) z_2. \end{aligned}$$

From (9) the entries of the strain tensor can be found and then by (10) the components of the stress tensor can be derived:

$$\begin{aligned} T_{11}(t) &= (-3 - 0.528e^{-16.804\Delta t} + 0.056e^{-11.12\Delta t} + 2e^{-8.402\Delta t} - 0.528e^{-5.436\Delta t} \\ &\quad + 2e^{-2.718\Delta t})\lambda + 2(-1.5 - 0.319e^{-16.804\Delta t} - 0.082e^{-11.12\Delta t} + 1.44e^{-8.402\Delta t} \\ &\quad - 0.099e^{-5.436\Delta t} + 0.56e^{-2.718\Delta t})\mu \\ T_{12}(t) &= 2(0.258e^{-16.804\Delta t} - 0.052e^{-11.12\Delta t} - 0.928e^{-8.402\Delta t} - 0.206e^{-5.436\Delta t} \\ &\quad + 0.928e^{-2.718\Delta t})\mu, \\ T_{21}(t) &= 2(0.258e^{-16.804\Delta t} - 0.052e^{-11.12\Delta t} - 0.928e^{-8.402\Delta t} - 0.206e^{-5.436\Delta t} \\ &\quad + 0.928e^{-2.718\Delta t})\mu, \\ T_{22}(t) &= (-3 - 0.528e^{-16.804\Delta t} + 0.056e^{-11.12\Delta t} + 2e^{-8.402\Delta t} - 0.528e^{-5.436\Delta t} \\ &\quad + 2e^{-2.718\Delta t})\lambda + 2(-1.5 - 0.209e^{-16.804\Delta t} - 0.137e^{-11.12\Delta t} \\ &\quad - 0.429e^{-5.436\Delta t} + 1.44e^{-2.718\Delta t})\mu. \end{aligned}$$

The velocity field is expressed via (5) by the equations

$$\dot{z}_1 = -6.81z_1 + 3.308z_2, \quad \dot{z}_2 = 1.969z_1 - 4.3z_2,$$

and the acceleration field is given through (8) by the equations

$$\ddot{z}_1 = 52.889z_1 - 36.787z_2, \quad \ddot{z}_2 = -21.892z_1 + 25.089z_2.$$

By substituting for $\mathbf{a}(\mathbf{z}, t)$ and $\mathbf{T}(\mathbf{z}, t)$ in Cauchy's equation of motion, (7), assuming the body forces to be equal to zero, we derive the system of the DEs

$$\begin{aligned} (52.889z_1 - 36.787z_2)\rho &= (-3 - 0.528e^{-16.804\Delta t} + 0.056e^{-11.12\Delta t} \\ &\quad + 2e^{-8.402\Delta t} - 0.528e^{-5.436\Delta t} + 2e^{-2.718\Delta t}) \frac{\partial \lambda}{\partial z_1} \end{aligned}$$

$$\begin{aligned}
&+2(-1.5 - 0.319e^{-16.804\Delta t} - 0.082e^{-11.12\Delta t} + 1.44e^{-8.402\Delta t} \\
&\quad - 0.099e^{-5.43648\Delta t} + 0.56e^{-2.718\Delta t}) \frac{\partial \mu}{\partial z_1} \\
&+2(0.258e^{-16.804\Delta t} - 0.052e^{-11.12\Delta t} - 0.928e^{-8.40176\Delta t} \\
&\quad - 0.206e^{-5.436\Delta t} + 0.928e^{-2.718\Delta t}) \frac{\partial \mu}{\partial z_2} \tag{17}
\end{aligned}$$

and

$$\begin{aligned}
(-21.893z_1 + 25.089z_2)\rho &= (-3 - 0.528e^{-16.804\Delta t} + 0.056e^{-11.12\Delta t} \\
&\quad + 2e^{-8.402\Delta t} - 0.528e^{-5.436\Delta t} + 2e^{-2.718\Delta t}) \frac{\partial \lambda}{\partial z_2} \\
&+2(0.258e^{-16.804\Delta t} - 0.052e^{-11.12\Delta t} - 0.928e^{-8.402\Delta t} \\
&\quad - 0.206e^{-5.436\Delta t} + 0.928e^{-2.718\Delta t}) \frac{\partial \mu}{\partial z_1} \\
&+2(-1.5 - 0.209e^{-16.804\Delta t} - 0.137e^{-11.12\Delta t} + 0.560e^{-8.402\Delta t} \\
&\quad - 0.426e^{-5.436\Delta t} + 1.44e^{-2.718\Delta t}) \frac{\partial \mu}{\partial z_2} \tag{18}
\end{aligned}$$

In order to solve the system of the DEs (17)-(18) we have to evaluate the density $\rho(t)$. Now, by assuming that the density depends only on the time -and not on the spatial coordinates- we get by the the continuity equation (16) that

$$\frac{\partial \rho}{\partial t} - 11.2\rho = 0. \tag{19}$$

5 The results

The numerical solution of the system of the DEs depends on the boundary conditions for the Lamé constants. By choosing suitable boundary conditions, i.e. big values for λ and μ in order to meet the exponential growth of the density given by (19), we get that the Lamé constants take still positive values as expected by the theory of real continua. The values of μ grow up very slowly as time increases, while the values of λ decrease slowly.

6 Conclusion

Since the features of a HMS do not give rise to the determination of certain boundary conditions concerning the evaluation of the Lamé constants, fixed numerical values can not be given to them. Nevertheless, by choosing some (random) boundary conditions for the Lamé constants, we can study their behavior as time increases. As noticed in section 5, it turns out that values of μ increase very slowly as time increases -in spite of the exponential growth of the

density. So it can be assumed that μ remains constant (at least over large time intervals). On the other hand the values of the λ coefficient decrease relatively very slowly as time increases, so it can be assumed that λ remains constant. The Young modulus and the Poisson factor remain also constant in the course of time, for the whole numerical grid. Since the aforementioned constants, appearing in the study of real elastic continua, retain their features while considered for the three-dimensional HMS-continuum, the evolution of this HMS can be interpreted as the deformation of a two-dimensional homogeneous elastic medium.

Acknowledgment: We would like to thank Prof. K. Soldatos for his contribution in formulating this problem.

References

- [1] Gani, J. (1963): Formulas for projecting enrolments and degrees awarded in universities, *J. R. Statist. Soc., A.* 135, 242-256.
- [2] Rogoff N. (1953): *Recent trends in occupational mobility*, Free Press, Glencoe, Illinois.
- [3] Patoucheas, P.D. and Stamou G. (1993): Non-homogeneous Markovian models in ecological modeling: a study of the zoobenthos dynamics in Thermaikos Gulf, Greece, *Ecological Modelling*, 66, 197-215.
- [4] Bartholomew, D.J. (1982): *Stochastic models for social processes*, third edn., Wiley, New York.
- [5] Davies, G. S. (1978): Attainable and maintainable regions in- Markov chain control, *Recent Theoret. Developm. in Control*, 371-381.
- [6] Vassiliou P.-C.G., Georgiou, A. C., and Tsantas, N. (1990): Control of asymptotic variability in non-homogeneous Markov systems, *J. Appl. Prob.*, Vol 27, 756-766.
- [7] Vassiliou, P.-C.G. (1984): Entropy as a measure of the experience distribution in a manpower system, *J. Operat. Res. Soc.*, Vol 35, 1021-1025.
- [8] Tsaklidis, G. (1998): The continuous time homogeneous Markov system with fixed size as a Newtonian fluid (?), *Appl. Stoch. Mod.and Data Anal.* Vol. 13; 177-182.
- [9] Tsaklidis, G. (1999): The stress tensor and the energy of a continuous time homogeneous Markov system with fixed size, *J.Appl. Prob.* Vol 36; 21-29.
- [10] Tsaklidis, G. and Soldatos, K. P. (2003): Modeling of continuous time homogeneous Markov system with fixed size as elastic solid: the two dimensional case, *Appl. Math. Modell.* Vol 27, 877-887.

Classification and Combining Models

Anabela Marques, Ana Sousa Ferreira and Margarida Cardoso

ESTBarreiro, Setúbal Polytechnic, Portugal, CEAUL

Email: anabela.marques@estbarreiro.ips.pt

LEAD, Faculty of Psychology, University of Lisbon, Portugal, CEAUL

Email: asferreira@fp.ul.pt

UNIDE, Dep. of Quantitative Methods of ISCTE - Lisbon University Institute, Portugal

Email: margarida.cardoso@iscte.pt

Abstract: In the context of Discrete Discriminant Analysis (DDA) the idea of combining models is present in a growing number of papers aiming to obtain more robust and more stable models. This seems to be a promising approach since it is known that different DDA models perform differently on different subjects. Furthermore, the idea of combining models is particularly relevant when the groups are not well separated, which often occurs in practice.

Recently, we proposed a new DDA approach which is based on a linear combination of the First-order Independence Model (FOIM) and the Dependence Trees Model (DTM). In the present work we apply this new approach to classify consumers of a Portuguese cultural institution. We specifically focus on the performance of alternative models' combinations assessing the error rate and the Huberty index in a test sample.

We use the R software for the algorithms' implementation and evaluation.

Keywords: Combining models, Dependence Trees model, Discrete Discriminant Analysis, First Order Independence model.

1. Introduction

Discrete Discriminant Analysis (DDA) is a multivariate data analysis technique that aims to classify and discriminate multivariate observations of discrete variables into *a priori* defined groups (a known data structure or Clustering Analysis results). Considering K exclusive groups G_1, G_2, \dots, G_K and a n -dimensional sample of multivariate observations - $X = (x_1, x_2, \dots, x_n)$ described by P discrete variables - DDA has two main goals:

1. To identify the variables that best differentiate the K groups;
2. To assign objects whose group membership is unknown to one of the K groups, by means of a classification rule.

In this work, we focus in the second goal and we consider objects characterized by qualitative variables (not necessarily binary) belonging to $K \geq 2$ *a priori* defined groups. We propose to use the combination of two DDA models: FOIM - First-

SMTDA 2010: Stochastic Modeling Techniques and Data Analysis
International Conference, *Chania, Crete, Greece, 8 - 11 June 2010*

Order Independence Model and DTM - Dependence Trees Model (DTM), Celeux
(1994) - to address classification problem.

In addition, we evaluate HIERM - Hierarchical Coupling Model performance when addressing the multiclass classification problems (Sousa Ferreira *et al.* (2000)) In order to evaluate the performance of the proposed approaches, we consider both simulated data and real data referred to consumers of a Portuguese cultural institution (Centro Cultural de Belém). Furthermore, we compare the obtained results with CART - Classification and Regression Trees (Breiman *et al.* (1984)) algorithm results.

2. Discrete Discriminant Analysis

The most commonly used classification rule is based on the Bayes's Theorem. It enables to determine the *a posteriori* probability of a new object being assigned to one of the *a priori* known groups. The Bayes's rule can be written as follows: if

$$\pi_l P_l(\underline{x}|G_l) \geq \pi_k P_k(\underline{x}|G_k) \text{ for } l=1, \dots, K \text{ and } l \neq k, (1)$$

then assign \underline{x} to G_k . π_l represents the *a priori* probability of group l (G_l), and $P(\underline{x}|G_l)$ denotes the conditional probability function for the l -th group. Usually, the conditional probability functions are unknown and estimated based on the training sample.

For discrete data, the most natural model is to assume that $P(\underline{x}|G_l)$ are multinomial probabilities estimated by the observed frequencies in the training sample, the well known Full Multinomial Model (FMM) (Celeux (1994)). However, since this model involves the estimation of many parameters, there are often related identifiability issues, even for moderate P . One way to deal with this high-dimensionality problem consists of reducing the number of parameters to be estimated recurring to alternative models proposals. One of the most important DDA models is the First-Order Independence Model (FOIM) (Celeux (1994)). It assumes that the P discrete variables are independent within each group G_k , the corresponding conditional probability being estimated by:

$$\hat{P}(\underline{x}|G_k) = \prod_{p=1}^P \frac{\#\{y \in G_k : y_p = x_p\}}{n_k} (2)$$

where n_k represents the G_k 's group sample dimension. This method is simple but is not realistic in some situations. Thus, some alternative models have been proposed. The Dependence Trees Model (DTM), Celeux (1994) and Pearl (1988), for example, takes the predictors' relations into account. In this model, one can estimate the conditional probability function, using a dependence tree that represents the most important predictors' relations. In this research, we use the

Chow and Liu algorithm (Celeux (1994) and Pearl (1988)) to implement the dependence tree and approximate the conditional probability function. In this algorithm, the mutual information $I(X_i, X_j)$ between two variables

$$I(X_i, X_j) = \sum_{x_i} \frac{\sum_{x_j} P(X_i, X_j) \log_2 P(X_i, X_j)}{P(X_i)P(X_j)} \quad (3)$$

is used to measure the closeness between two probability distributions. For example, take $P = 4$ variables and consider the data listed in Table 2. For each pair of variables we obtain the mutual information (see Table 1). Since $I(x_2, x_3)$, $I(x_1, x_2)$ and $I(x_2, x_4)$ correspond to the three largest values the branches of the best dependence tree are (x_2, x_3) , (x_1, x_2) and (x_2, x_4) and

$$P^*(x|G_k) = P(x_1|2)P(x_3|x_1,2)P(x_2|x_1,1)P(x_4|x_1,2) \quad (4)$$

Table 2 illustrate the differences between the estimates of the 3 referred DDA models corresponding to the data considered. Note that the DTM model estimates are closer to the FMM estimates but there are no null frequencies.

| (x_i, x_j) | $I(x_i, x_j)$ |
|--------------|-----------------|
| (x_1, x_2) | 0,079434 |
| (x_1, x_3) | 0,000051 |
| (x_1, x_4) | 0,005059 |
| (x_2, x_3) | 0,188994 |
| (x_2, x_4) | 0,005059 |
| (x_3, x_4) | -0,024 |

Table 1. Mutual information values

| (x_1, x_2, x_3, x_4) values | num. observ./ G_k | $\hat{P}(x_1, x_2, x_3, x_4)$ for group G_k | | |
|-------------------------------|---------------------|---|------|------|
| | | FMM | FOIM | DTM |
| 0000 | 10 | 0,10 | 0,05 | 0,10 |
| 0001 | 10 | 0,10 | 0,05 | 0,13 |
| 0010 | 5 | 0,05 | 0,06 | 0,03 |
| 0011 | 5 | 0,05 | 0,06 | 0,04 |
| 0100 | 0 | 0,00 | 0,06 | 0,02 |
| 0101 | 0 | 0,00 | 0,06 | 0,02 |
| 0110 | 10 | 0,10 | 0,07 | 0,08 |

| | | | | |
|------|----|------|------|------|
| 0111 | 5 | 0,05 | 0,07 | 0,07 |
| 1000 | 5 | 0,05 | 0,06 | 0,04 |
| 1001 | 10 | 0,10 | 0,06 | 0,05 |
| 1010 | 0 | 0,00 | 0,07 | 0,01 |
| 1011 | 0 | 0,00 | 0,07 | 0,02 |
| 1100 | 5 | 0,05 | 0,07 | 0,04 |
| 1101 | 5 | 0,05 | 0,07 | 0,03 |
| 1110 | 15 | 0,15 | 0,08 | 0,18 |
| 1111 | 15 | 0,15 | 0,08 | 0,15 |

Table 2. Conditional probability estimates for group G_k

3. Combining Models in Discrete Discriminant Analysis

The idea of combining models currently appears in an increasing number of papers. The aim of this strategy is to obtain more robust and stable models. Sousa Ferreira (2000) proposes combining FMM and FOIM to address classification problems with binary predictors, using a single coefficient β ($0 \leq \beta \leq 1$) to weight these models. In spite of yielding good results, the referred approach shows that the resulting FMM weights tend to be frequently negligible, even when the observed frequencies are smoothed (Brito *et al.* (2006)).

In view of these conclusions, Marques *et al.* (2008) propose a new model which has an intermediate position between the FOIM and DTM models:

$$\hat{F}(x|\beta) = \beta \hat{F}_{FOIM}(x) + (1 - \beta) \hat{F}_{DTM}(x) \quad (5)$$

In the present work the combining models' parameter is assigned different values ranging from 0 to 1.

4. The Hierarchical Coupling Model

In the multiclass case ($K \geq 2$) we can recur to the Hierarchical Coupling Model (HIERM) (Sousa Ferreira *et al.* (2000)) that decomposes the multiclass problem into several biclass problems using a binary tree structure. It implements two decisions at each level:

1. Binary branching criterion for selecting among the possible 2^{K-1} -1 groups combinations;
2. Choice of the model or combining model that gives the best classification rule for the chosen couple.

In the present work we implement branching using the affinity coefficient, Matusita (1955) and Bacelar-Nicolau (1985). Supposing $F_1=\{p_i\}$ and $F_2=\{q_i\}$, $i=1,\dots,L$ are two discrete distributions defined on the same space, the correspondent affinity coefficient is computed as follows:

$$\rho(F_1, F_2) = \sum_{i=1}^L \sqrt{p_i} \sqrt{q_i} \quad (6)$$

The process stops when a decomposition of groups leads to single groups. For each biclass problem we consider FOIM, DTM or an intermediate position between them.

5. Numerical Experiments

We conduct numerical experiments for simulated data and real data using moderate and large samples, respectively. We use test samples to evaluate the alternative models precision. Indicators of precision are the percentage of correctly classified observations (P_c) and the Huberty index:

$$\frac{P_c - P_d}{1 - P_d}$$

where P_d represents the percentage of correctly classified cases using the majority class rule.

5.1 Simulated data

The simulated data set considered has 250 observations, 4 groups and 3 binary predictors (see Table 3). To evaluate the proposed models' performance we use precision corresponding to a test (sub)sample: 50% of the original sample. The modal class in the test sample has 32% of the observations which yields the same 32% for percentage of correctly classified observations, according to the majority rule.

| | Total data set | | Training sample | | Test sample | |
|-------|----------------|-----|-----------------|-----|-------------|------------|
| | n_k | % | n_k | % | n_k | % |
| G_1 | 80 | 32% | 40 | 32% | 40 | 32% |
| G_2 | 70 | 28% | 35 | 28% | 35 | 28% |
| G_3 | 30 | 12% | 15 | 12% | 15 | 12% |
| G_4 | 70 | 28% | 35 | 28% | 35 | 28% |

Table 3. Characterization of simulated data set

The results obtained are presented in Table 4. For this data set the HIERM model and FOIM-DTM combination yield the best results.

| Classification Method | | % of correctly classified | Huberty index | |
|-------------------------------------|-------------------------------------|---------------------------|---------------|---------------|
| CART | | 45,6% | 20,00% | |
| | $\beta = 0$ (DTM) | 52,8% | 30,59% | |
| β *FOIM+ (1- β)*DTM | $\beta = 0,25$ | 47,2% | 22,35% | |
| | $\beta = 0,50$ | 48,8% | 24,71% | |
| | $\beta = 0,75$ | 48,8% | 24,71% | |
| | $\beta = 1$ (FOIM) | 48,8% | 24,71% | |
| MHIERM: G_2+G_1 vs G_3+G_4 | $\beta = 0$ (DTM) | 45,6% | 20,00% | |
| | $\beta = 0,25$ | 59,2% | 40,00% | |
| | $\beta = 0,50$ | 60,8% | 42,35% | |
| | β *FOIM+ (1- β)*DTM | $\beta = 0,75$ | 60,8% | 42,35% |
| | $\beta = 1$ (FOIM) | 59,2% | 40,00% | |

Table 4. Simulated data set results

5.2 Real data

We consider a data set referred to 988 observations originated from questionnaires directed to consumers of Centro Cultural de Belém, a Portuguese cultural institution (Duarte (2009)). Data includes three questions related to the quality of services provided by CCB that this study tries to relate with consumers' education: we specifically use 4 education levels as the target variable. Predictors are: X_1 - Considering your expectations how do you evaluate CCB products and services?(1=much worse than expected ...5=much better than expected); X_2 - How do you evaluate CCB global quality? (1=very bad quality ...5=very good quality); X_3 : How do you evaluate the price/quality relationship in CCB?(1=very bad...5=very good). The groups distribution is illustrated in Table 5.

| | Total data set | | Training sample | | Test sample | |
|-------|----------------|-----|-----------------|-----|-------------|-----|
| | n_k | % | n_k | % | n_k | % |
| G_1 | 177 | 18% | 115 | 18% | 62 | 18% |
| G_2 | 136 | 14% | 88 | 14% | 48 | 14% |

| | | | | | | |
|-------|-----|-----|-----|-----|-----|-----|
| G_3 | 462 | 47% | 300 | 47% | 162 | 47% |
| G_4 | 213 | 22% | 138 | 22% | 75 | 22% |

Table 5. Characterization of CCB data set

The results obtained are presented in Table 6. For CCB problem the best solution is achieved by HIERM model and combined FOIM-DTM model.

| Classification Method | | % of correctly classified | Huberty index |
|------------------------|--------------------|---------------------------|---------------|
| CART | | 46,10% | -1,70% |
| β *FOIM+ | $\beta = 0$ (DTM) | 45,00% | -3,77% |
| (1- β)*DTM | $\beta = 0,20$ | 45,80% | -2,26% |
| | $\beta = 0,40$ | 46,40% | -1,13% |
| | $\beta = 0,50$ | 47,60% | 1,13% |
| | $\beta = 0,60$ | 47,30% | 0,57% |
| | $\beta = 0,80$ | 47,80% | 1,51% |
| | $\beta = 1$ (FOIM) | 47,00% | 0,00% |
| MHIERM: | $\beta = 0$ (DTM) | 47,80% | 1,51% |
| G_2 vs $G_1+G_3+G_4$ | $\beta = 0,20$ | 48,10% | 2,08% |
| | $\beta = 0,40$ | 49,30% | 4,34% |
| β *FOIM+ | $\beta = 0,50$ | 49,30% | 4,34% |
| (1- β)*DTM | $\beta = 0,60$ | 49,30% | 4,34% |

| | | |
|--------------------|---------------|--------------|
| $\beta = 0,80$ | 48,40% | 2,64% |
| $\beta = 1$ (FOIM) | 49,90% | 5,47% |

Table 6. CCB data set results (test sample)

6. Conclusions and perspectives

In the present work we propose using the combination of two DDA models (FOIM and DTM) and use the HIERM algorithm to address classification problems. We compare results obtained with CART results into two data sets: simulated data (250 observations) and real data (988 observations). We use indicators of classification precision obtained in the test set (we consider 50% and 35% of observations for the smaller and larger data set, respectively).

According to the obtained results, the proposed approach performs slightly better than CART, on both simulated and real data. However, the classification precision attained barely attains the precision corresponding to the majority class rule in the real data set. In fact, in this case we are dealing with very sparse data (46% of the multinomial cells have no observed data in any of the groups considered) which turns the classification task very difficult.

In future research, the number of numerical experiments should be increased, both using real and simulated data sets and considering several sample dimensions. The number of variables considered (binary and non-binary) should not originate an excessive number of states (around a thousand) due to the number of parameters that need to be estimated. Alternative strategies to estimate the β parameter, such as least squares regression, likelihood ratio or committee of methods, should also be considered.

References

1. Bacelar-Nicolau, H., The Affinity Coefficient in Cluster Analysis, in *Meth. Oper. Res.*, **53**: 507-512 (1985).
2. Breiman, L., Friedman, J.H., Olshen, R. A. and Stone, C.J., *Classification and Regression Trees*, Wadsworth, Inc. California (1984).

3. Brito, I., Celeux, C. and Sousa Ferreira, A., Combining Methods in Supervised Classification: a Comparative Study on Discrete and Continuous Problems. *Revstat – Statistical Journal*, Vol. 4(3), 201-225 (2006).
4. Celeux, G. and Nakache, J. P., Analyse Discriminante sur Variables Qualitatives. G. Celeux et J. P. Nakache Éditeurs, *Polytechnica*, (1994).
5. Duarte, A., *A satisfação do consumidor nas instituições culturais. O caso do Centro Cultural de Belém*. Master Thesis. ISCTE, Lisboa (2009)
6. Marques, A.; Sousa Ferreira, A. and Cardoso, M. Uma proposta de combinação de modelos em Análise Discriminante. *Estatística – Arte de Explicar o Acaso*, in Oliveira, I. et al. Editores, *Ciência Estatística*, Edições S.P.E, 393-403 (2008).
7. Matusita, K., Decision rules based on distance for problems of fit, two samples and estimation. In *Ann. Inst. Stat. Math.*, Vol. 26(4): 631-640, (1955).
8. Pearl J., *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Los Altos: Morgan Kaufmann, (1988).
9. Sousa Ferreira, A., *Combinação de Modelos em Análise Discriminante sobre Variáveis Qualitativas*. PhD thesis. Universidade Nova de Lisboa, (2000).
10. Sousa Ferreira, A.; Celeux, G. and Bacelar-Nicolau, H., Discrete Discriminant Analysis: The performance of Combining Models by an Hierarchical Coupling Approach. In Kiers, Rasson, Groenen and Shader, editors, *Data Analysis, Classification and Related Methods*, pages 181-186. Springer, (2000).