HAL
open science

Finite sampling design with quantization Applied to test of the bioavaibility

Kamal Boukhetala, Karim Benhenni, Zohra Habib

To cite this version:

Kamal Boukhetala, Karim Benhenni, Zohra Habib. Finite sampling design with quantization Applied to test of the bioavaibility. 2011. hal-00619141

HAL Id: hal-00619141

https://hal.science/hal-00619141v1

Preprint submitted on 5 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Finite sampling design with quantization
Applied to test of the bioavaibility

K.Boukhetala & Z.Habib (*)
(*) Département de Recherche Opérationnelle
Faculté des Sciences Mathématiques
Bp.32, U.S.T.H.B, El Alia, Bab-Ezzouar
Alger, Algérie

K. Benhenni (**)
(**) Laboratoire de Statistique et d'Analyse de Données de Grenoble Bt. Sciences Humaines et Mathématiques Université II de Grenoble,
France.

Abstract:

The optimal sampling problem for the estimation of an integral, based on observations, with random and correlated measurement errors is treated. Considering the effect of the quantization of errors, we determine the optimal design, which is constructed by utilizing stochastic algorithms. The efficiency of the Genetic Algorithm is confirmed by comparison with other implemented algorithms, by considering a pharmacokinetic application for a test of bioavailability. A multiobjective approach is proposed for the choice of an optimal compromise design between the variability of the quantization of errors and the integral's estimator..

Keywords. Bioavailability, Optimal sampling, Quantization, Genetic Algorithm, Annealing Genetic Algorithm, Multiobjective Evolutionary Algorithms.

0. INTRODUCTION

In many experimental situations, such as the study of the bioavailability: a particular problem of pharmacokinetic, we are led to study the effect of the amount of drug absorbed through the drug's concentration over a time interval [a,b] represented by the area under a curve which it could be computed based on a quadrature estimator constructed from some appropriate observations with random measurement errors. It's known that the concentration function depends very much on the way of the drug's administration : oral or intravenous, so we try to determine for both cases the optimal sampling design of finite observations relatively to the time interval [a,b]. The estimator is constructed from this optimal design using a quadrature rule of the integral approximation such as the trapezoidal rule from fixed sample size n. The performance of this estimator is measured by the mean squared error of the approximation between the estimator and the area under the concentration curve for any fixed sample size n.

Darryl Katz and David. Z D'argenio, [1] studied this problem when the measurement errors are independent, with the Gaussian distribution. They implemented an algorithm which minimized the expectation of the square of the difference between the exact integral and the quadrature approximation with the aim of finding the optimal sampling design.

(Benhenni and Cambanis (1992), [2]) were able to find an asymptotically (as n tends to infinity) optimal design for the problem of estimating an integral of a stochastic process. Also, in order to ameliorate the performance of the sampling design, (Benhenni and Cambanis (1998), [3]) have shown that the quantization reduces the convergence's rate. In (Boukhetala, Benhenni and Benamara, (1996), [4]), it is supposed that, in the problem treated by Katz (1983), the measurement errors are no longer independent but they are correlated and belong to a particular class of Gaussian process.

The assumption of correlated errors makes the optimality criteria more complicated and less stable. The analytical approach to solve this problem is difficult for fixed sample size.

So, in addition to a mathematical method, we propose a Genetic Algorithm (AG) and we compare its performance with other known "heuristic" algorithms, such as the genetic annealing algorithm.

In this paper, we add the effect of the quantization on measurement errors and we show that for a fixed sample, the results are ameliorated. In order to give more choices to the decision maker, we propose a multiobjective approach to resolve the problem, thus two Multiobjective evolutionary algorithms adapted to the problem are implemented.

1. ONE DIMENSIONAL ASYMPTOTIC SAMPLING RESULTS

In general, sampling design aims at an arrangement of observations of distributed natural process under certain criterion of exploration. In many situations, we seek to minimize the estimating error. The variability characteristics of the natural process sampled must to be considered by the sampling technique. To concept an efficient sampling , it necessary to effect an analysis of variability. The optimal sampling designs, in a certain mathematical sense is a difficult problem; in one dimension (see, for example, K.Benhenni et S.Cambanis, [3]). In [3], the problem of the determination of an asymptotic

optimal sampling for the estimation of the integral I(X)= $\int_0^1 X(t)\phi(t)dt$, is considered. X(t) is a (measurable) process with mean 0 and continuous covariance function R(s,t)=E(X(s)X(t)) that have exactly K derivatives. The weight ϕ is a known (nonrandom) function (in $L_2[0,1]$).

The trapezoidal predictor used to estimate the random integral I(X) is $I_n = \sum_{i=0}^{n} c_{i,n} X(t_{i,n})$ and the aim is to find an asymptotic optimal designs $(T^*_n)_n$ of a regular sequences of sampling designs $(T_n(h)=\{T_{i,n}\}^n_{i=O}$, $\{0=t_{0,n} < t_{1,n} < \ldots t_{n,n}=1\}_{n=1}^{\infty}$ generated by a positive continuous density h via $\int_0^{t_{i,n}} h(t)dt = \frac{i}{n},\ n = 0,1,\ldots n,$ and estimator-coefficients $(C^*_n)_n$, satisfying:

$$\frac{E(I - I_n^*)^2}{\inf_{T_n,\, C_n} E(I - I_n)^2} \to_n 1$$

The following theorem shows the asymptotic performance under appropriate regularly conditions.

Theorem (Benhenni, Cambanis, [3]) :

Under a certain satisfied assumptions ,

$$n^{2K+2} E(I - g_n{}' R^{-1}{}_n X_n)^2 \to_n \frac{|B_{2K+2}|}{(2K+2)!} \int_0^1 \frac{\phi^2(t)\alpha_K(t)}{h^{2K+2}(t)} dt .$$

Where

- $R_n=\{R(t_{i,n},t_{j,n})^n_{i,j=0}$, B_m is the m-th Bernouli number,
- $\alpha_K(t) = R^{(K,K+1)}(t,t-) - R^{(K,K+1)}(t,t+)\ (\geq 0)$
- $g'_n=(g(t_{0,n}),\ldots,g(t_{n,n})$, with $g(t) = \int_0^1 R(s,t)\phi(s)ds$.

We steady the special case: X(t)= f(t) + $\varepsilon(t)$, where $\varepsilon(t)$ is a Gaussian process with E($\varepsilon(t)) = 0$ and the Function f(t) is deterministic, defined in [0,1],.

For the estimator $I_n(X) = \frac{1}{n}\sum_{i=1}^{n} a_i X(t_i)$ of the integral I(f)= $\int_0^1 f(t)dt$, we show the following result, that gives the performance of the sampling design.

Proposition:

We have

1- $\sigma^2{}_n = E(I(f) - I_n(X))^2 \to_n \sigma_\infty{}^2 = R(t,s) = \int_0^1\int_0^1 E(\varepsilon(t)\varepsilon(s)dtds$

2- $n^2(\sigma_n{}^2 - \sigma^2{}_\infty) \to_n C$

Proof

1)

We have

$$\sigma^2{}_n = E(I(f) - I_n(X))^2 = E(I(f) - I_n(f) - I_n(\varepsilon))^2$$

$$= (I(f) - I_n(f))^2 + E((I_n(\varepsilon))^2 \to_n \sigma_\infty{}^2 = R(t,s) = \int_0^1\int_0^1 E(\varepsilon(t)\varepsilon(s)dtds$$

For the point 2),

$$(\sigma_n{}^2 - \sigma^2{}_\infty) = (I(f) - I_n(f)^2 + E(I_n(\varepsilon))^2 - E(I(\varepsilon))^2;$$

$$E(I_n(\varepsilon))^2 - E(I(\varepsilon))^2 =$$

$$\frac{1}{4n^2}\sum_i\sum_j\{R(t_i,t_j) + R(t_i,t_{j+1}) + R(t_{i+1},t_j) + R(t_{i+1},t_{j+1})\ \} - \int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} R(t,s)dtds.$$

$$= \sum_i\sum_j M_{ij}$$

Where for t>s, we have

$$M_{ii} = \frac{1}{4n^2}\left[R(t_i,t_i) + 2R(t_i,t_{i+1}) + R(t_{i+1},t_{i+1})\right] - 2\int_{t_i}\int_{t_i} R(t,s)dtds$$

$R(t,s)=R(t,t_i)+(s-t_i)R^{(0,1)}(t+,\xi_s)=R(t_i,t_i)+(t-t_i)R^{(1,0)}(\eta_{i^+},t_i)+(s-t_i)R^{(0,1)}(t+,\xi_s)$, for $t_i < \xi_s < s$

However

$$\int_{t_i}^{t_{i+1}} dt \int_{t_i}^{t} R(s,t)ds = \begin{cases} R(t_i,t_i)\left\{\int_{t_i}^{t_i+1} dt(t-t_i) = \frac{(\Delta t_i)^2}{2}\right\} \\ + R^{(1,0)}(\eta_i,t_i)\left\{\int_{t_i}^{t_{i+1}} (t-t_i)^2\, dt = \frac{(\Delta t_i)^3}{3}\right\} \\ + R^{(0,1)}(t_i,t^-{}_i)\left\{\int_{t_i}^{t_{i+1}} dt \int_{t_i}^{t} (s-t_i) = \frac{(\Delta t_i)^3}{6}\right\} \end{cases}$$

$R(t_i,t_{i+1}) = R(t_i,t_i) + \Delta t_i\, R(t_i,t+_i)$.

$R(t_{i+1},t_{i+1}) = R(t_i,t_{i+1}) + \Delta t_i\, R^{(0,1)}(t_i,\xi_i) = R(t_i,t_i) + \Delta t_i\, R^{(1,0)}(t+_i,t_i)) + \Delta t_i\, R^{(0,1)}(t_{i+1},\xi_i)$.

So

$$M_{ii} = -\frac{1}{12n^3}(R(0,1)(t,t-) - R(0,1),t,t+) = -\frac{1}{12}\alpha(t)$$

And

$$n^2(E((I_n{}^2(\varepsilon) - E(I^2(\varepsilon)) \to_n -\frac{1}{12}\int_0^1 \frac{\alpha(t)}{h^2(t)}dt$$

Thus, $(\sigma_n{}^2 - \sigma^2{}_\infty) \to_{n\to\infty} \frac{C}{n^2}$.

1. FINITE SAMPLING DESIGN WITH QUANTIZATION APPLIED TO TEST THE BIOAVAIBILITY OF MEDICAMENT

For many real problems, it is interesting to consider a finite sampling design. The above theoretical results on the performance of the sampling design, motive us to treat a important problem, that consist to test the bioavaibility of medicament. We seek to determine an optimal finite sampling design on the time of the process, by using various approaches to solve the problem. The various proposed approaches required to apply a quantization process, to accelerate the speed of algorithms associated these approaches, in particular the genetic algorithms. The case without quantization is treated in [.].

2.1. Modeling (K. Boukhetala, Z. Habib, [7])

In order to study the bioavailability, Darryl Katz and David Z. D'Argenio [1] have considered the following problem : Let f a real function , which expresses the drug concentration over time interval [0,T]. In practical situations, the effect of concentration is observed at times $\{t_0,t_1,\ldots\ldots t_{m+1}\}$ (t_0 and t_{m+1} are fixed. The observations h_i of f, at time t_i are supposed accompanied of measurement errors ε_i ; random independent Gaussian variable, with null mean and variance equalize to σ^2_i : $h_i = f(t_i) + \varepsilon_i$ i=0,1……m+1.
The problem consists in seeking an optimal design $\{t_1,\ldots\ldots t_m\}$, by approximating the integral:

$$I = \int_{t_0}^{t_{m+1}} f(t)dt \ ,$$

The trapezoidal estimator is given by : $T = \frac{1}{2}\left\{h_0(t_1 - t_0) + \sum_{i=1}^{m} h_i(t_{i+1} - t_{i-1}) + h_{m+1}(t_{m+1} - t_m)\right\}$

By applying a quantization function Q(.) on the errors process , the mathematical problem to resolve is:

$$(P)\begin{cases} Min \ \ (J = E(T - I)^2) \\ h_i = f(t_i) + Q(\varepsilon_i), \quad i = 1 \ldots\ldots m \\ t_i \le t_{i+1} \end{cases}$$

A general quantizer approximates an observed value by the nearest among a finite number of representative values.
The simplest and most common form of quantizers is the scalar quantizer. A vector quantizer Q of dimension k and size N is formally a mapping from k-dimensional Euclidean space R^k into a finite set C, containing N output or reproduction vectors from R^k. For a choice of an efficient random quantizer, it is interesting to consider E[d(x,Q(x)] to minimize, where E represents the expectation operator and d is a given distance.
Under the assumption of the dependence of the errors and their quantization, the objective function becomes

$$\begin{aligned} J = {}& \tfrac{1}{4}\left\{\sigma_0^2(t_1-t_0)^2 + \sum_{i=1}^{m}\sigma_i^2(t_{i+1}-t_{i-1})^2 + \sigma_{m+1}^2(t_{m+1}-t_m)^2\right\} \\ & + \tfrac{1}{2}\{\ \sigma_0(t_1-t_0)\sigma_{m+1}(t_{m+1}-t_m)\rho(t_0,t_{m+1}) + \sigma_0(t_1-t_0)\sum_{i=1}^{m}\sigma_i(t_{i+1}-t_{i-1})\rho(t_0,t_i) \qquad \rho(t_i,t_j) = E(Q(\varepsilon_i)Q(\varepsilon_j)) \\ & + \sigma_{m+1}(t_{m+1}-t_m)\sum_{i=1}^{m}\sigma_i(t_{i+1}-t_{i-1})\rho(t_i,t_{m+1}) + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\sigma_i(t_{i+1}-t_{i-1})\sigma_j(t_{j+1}-t_{j-1})\rho(t_i,t_j)\ \} \\ & + \left\{\tfrac{1}{2}\left\{f_0(t_1-t_0) + \sum_{i=1}^{m} f_i(t_{i+1}-t_{i-1}) + f_{m+1}(t_{m+1}-t_m)\right\} - I\right\}^2. \end{aligned}$$

and

$$\sigma_i = \sqrt{Var\ \ (Q(\varepsilon_i))}$$

3 . APPRAOCHES USED TO SOLVE THE PROBLEM

In our case, finding an exact analytic solution is a difficult problem, because of the complicated form of the gradient of the objective function (we expressed it above). For the resolution of the problem we proposed two heuristics that we judged interesting for our problem. We showed among many conclusions that the Genetic Algorithm is the most effective for this problem. We used the quantization to improve more again the performance of methods used for the research of optimal sampling, in particular the genetic algorithm. We would mention that it has already been proven asymptotically by Karim Benhenni and Stamatis Cambanis [3] that the rate of convergence reduces considerably when the sample is quantified. We used the approach of multi- objective optimization to allow the decision maker to have several choices.
We present in the following paragraph the several methods used, and the results obtained by their application to the pharmaceutical problem.

3.1. The Mathematical Approach

We showed that the use of this method is computationally very expensive which is not interesting in our case. That is surely due to the form of the gradient which is complicated

3.2. Genetic Algorithm Approach (AG)

A Genetic algorithm is an optimization tool which mimics the evolution of a population towards fitness to the natural environment. It is an iterative procedure that use in general three evolutionary operators: reproduction, crossover and mutation. Each iteration of an AG yields a new population from an existing one.
On the basis of these considerations, we have developed an AG , adapted to the treated problem.

3.3. Genetic Annealing Algorithm Approach (RG)

The Simulated Annealing Algorithm (SAA) and the Genetic Algorithm (AG) have proved separately their effectiveness for many practical problems. The RG is an hybrid method which is a combination of SAA and AG. This method explores the advantages of the two algorithms in the hope to improve the solution.. The RG was implemented to solve the problem (P) and its results were compared to those of the other algorithms considered, as it will be indicated by the table 1.

3.4. Multi-criteria Approach

The principles of multi-criterion optimization are different from the ones of a single-objective optimization.
The main goal in a single-objective optimization is to find the global optimal solution. However, in a multi-criterion optimization problem, there are more than one objective function, each of then may have a different individual optimal solution. If there is sufficient difference in the optimal solutions corresponding to different objectives, the objective functions are often known as conflicting to each other. Multi-criterion optimization with such conflicting objective functions gives rise to a set of optimal solutions, instead of one optimal solution.
Evolutionary algorithms seem particularly desirable to solve multi-objective optimization problems because they deal simultaneously with a set of possible solutions (the so-called population) which allows to find an entire set of Pareto optimal solutions in a single run of the algorithm. In order to solve our problem by using this approach, we chose the implementation of two methods : The (N.Srinivas and Kayannoy Deb's, 2000, [5) non-dominated sorting genetic algorithm (NSGA)] and Multi-Sexual Genetic Algorithm (Joanna Lis and A.E.Eiben, 1996, [6]) .
In our problem, the two criteria taken into account are respectively a component related to the observation error and a component based on the quadrature error in approximating the integral of the exact function.

4. NUMERICAL RESULTS AND COMMENTS

The various approaches quoted previously were tested for the calculation of the optimal sampling. In this context, the application concerns the bioavailability test.
Two models proposed by Darryl Katz and D.Z.D'Argenio were chosen for the application.

Model 1 " The oral administration ":

In this case, we consider the following three-exponential model which represents the plasma concentration (ng/ml) of the drug digoxin:
$f_1(t)=A\exp(-\lambda_1 t)+B\exp(-\lambda_2 t)+C\exp(-\lambda_3 t)$, where
$A = -2{,}4$; $\lambda_1 = 1\ 0$; $B = 2$; $\lambda_2 = 0{,}65$; $C = 0{,}4$
and $\lambda_3 = 0{,}0146$.

Model 2 " The intravenous administration":

The intravenous concentration measurements were generated by using the following function:
$f_2(t)=D\exp(-\lambda_2 t)+E\exp(-\lambda_3 t)$, where
$D=3{,}117$; $E=0{,}6657$; $\lambda_2=0{,}65$; $\lambda_3=0{,}0146$
$t \in [0,\ 96\ \ hours\ ]$.
The following function is assumed to be the standard deviation of the observation (assay) error :
$\sigma_i = 0.05+0.1f(t_i)$.
The correlated measurement errors process is supposed to be gaussian and belongs to the class G={Analytic (AN), Stationary Gauss-Markov (SGM)}.
The process AN and SGM are characterized respectively by the correlation functions, given by:
$\rho_{AN}(t_i,t_j)=e^{-\alpha|t_i-t_j|}$ and $\rho_{SGM}(t_i,t_j)=e^{-\alpha(t_i-t_j)^2}$ respectively. α is a parameter estimated by simulations.
With a suitable choice of the parameters of the implemented methods, and in the case of independent Gaussian measurement errors, we give by the following graphical representation a comparative study between AG and RG methods. The comparative criteria are the evolution of the function J (criteria 1), the CPU time (criteria 2) and the accurcy (criteria 3). During the experiments, several sizes of samples were taken into account. The graphical results concerning these criteria are given by the figure 1.
We noticed that the method RG was distinguished from AG when we increased the size of the problem.

Let us note the fact that after several simulations, we noted the inefficiency of the mathematical method as we suspected it because of the complexity of the gradient. Throughout the various tests, we noted the effectiveness of the genetic algorithms which confirmed the results found previously.
By using the quantization, we improved broadly the results obtained before. After many simulations, we obtained the interval variations of the amelioration rate: Scalar Quantization: [1%, 28%], Vectorial Quantization: [1%, 31%].
The objective function J breaks up into J_1 and J_2 which are respectively : a component related to the observation error and a component based on the quadrature error in approximating the integral.
Finally, we implemented two multicriterion methods to try to help of advantage the decision maker by taking account of the two criteria J_1 and J_2.
Being given that it is not possible to present all the solutions of Pareto obtained with NSGA and MSGA within the framework of this article, we will present just the solution minimizing J_1+J_2 through two examples.
We will note $\mathbf{J_1^*}$ and $\mathbf{J_2^*}$ respectively the optimum values of the functions objectives obtained with AG.
By carrying out various simulations, we confirmed the effectiveness of the genetic algorithms for obtaining the optimal sampling for the estimation of the bioavailability. The quantization gives more precise results in general and accelerates the process of search of the optimal solution. By bringing more flexibility, the multicriteria approach is a new approach which helps the decision maker.

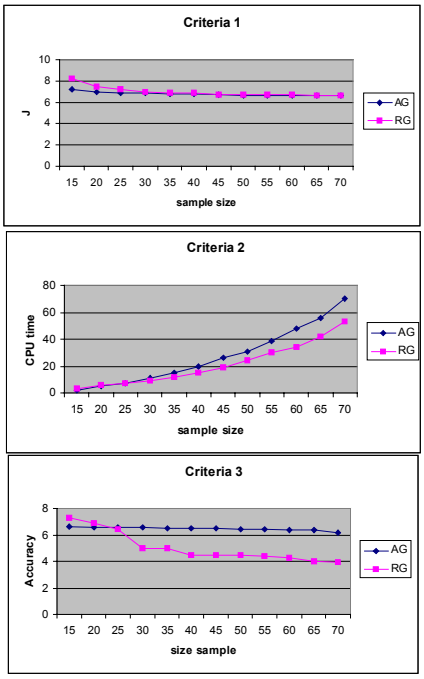
Table 1: Results of two examples obtained by the NSGA and MSGA approaches for SGM and AN models

Method	Model	α^*	J_1^*	J^*_2	J_1	J_2
NSGA	GMS	0.4	3.8	6.66 E^{-9}	5,24	0,001
NSGA	AN	*0.1*	*3.5*	*5.1 E^{-7}*	*5,02*	*0,01*
MSGA	GMS	*0.4*	*3.8*	*6.66 E^{-9}*	*4,3*	*0,0001*
MSGA	AN	*0.1*	*3.5*	*5.1 E^{-7}*	*5,32*	*0,07*

(J_1, J_2) represents the solution of Pareto.

The graphical results concerning these criteria are given by the following figures

Figure 1: Comparison between algorithms AG and RG, for the three criterion



5. IMPLEMENTATION

We have implemented the different methods of resolution on a personal computer, by using the Borland Builder C++ language. The package was developed under the windows environment which brings to the user an easy and a flexible tool to realize many simulations for the calculation of the bioavailability before each laboratory experimental.

6. CONCLUSION

By carrying out various simulations, we confirmed the effectiveness of the genetic algorithms for obtaining the optimal sampling for the estimate of the bioavailability. The quantization gives more precise results in general and accelerates the process of search of the optimal solution. By bringing more flexibility, the multicriteria approach is a new approach which helps the decision maker. The developed software makes it possible to the user to carry out himself various experimental tests of bioavailability by carrying out simulations, this in particular makes it possible to reduce the risk of medication and allows to control the parameters of the problem.

REFERENCES

[1] Darryl Katz & David. Z D'argenio (1983). Experimental Design for estimating Integral by numerical quadrature, with applications to pharmacokinetic studies. *Biometrics*. **39**, pp 621-628.

[2] Benhenni and Cambanis (1992). Sampling Design for estimating integrals of stochastic process. *The annals of statistics*. **20**, pp 161-194.

[3] Benhenni and Cambanis (1998). The effect of quantization on the performance of sampling design. *IEEE transactions on information theory*, **44**, N°5.

[4] Boukhetala, Benhenni and Benamara, (1996). Optimal sampling estimating integral of function from observations with correlated measurement errors, *Compstat'96,* **2**, pp 161-163.

[5] N.Srinivas & Kalyannoy D. (2000). Multiobjective optimization using nondomined sorting in GA. *Departement of mecanical engineering* .
http://www.Lania.mx/ncceollo/emmo/emoobib.html

[6] Lis A.E Joanna (1996). Multi-sexual genetic algorithm for multiobjective optimizaion. Eiben. *Proceedings of the International Conference on evolutionary Computation (IEEE), Japan*.
http://www.Lania.mx/ncceollo/emmo/emoobib.html.

[7] K. Boukhetala and Z. Habib (2002) : Finite sampling design with quantization for a pharmacokinetic problem:
Proceedings of the 6th World Multiconference Systemics, Cybernetics and Informatic, Orlando, Florida.