



ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

C. R. Acad. Sci. Paris, Ser. I

[www.sciencedirect.com](http://www.sciencedirect.com)

Statistique

## Plans d'expérience pour l'estimation de la courbe de concentration et de l'AUC

*Experimental designs for the estimation of the concentration curve and the AUC*

Mohamad Belouni<sup>a</sup>, Karim Benhenni<sup>b</sup>

<sup>a</sup> Laboratoire Jean-Kuntzmann (UMR CNRS 5224), université Joseph-Fourier, BP 53, 38041 Grenoble cedex 09, France

<sup>b</sup> Laboratoire Jean-Kuntzmann (UMR CNRS 5224), université Pierre-Mendes-France, BP 47, 38040 Grenoble cedex 09, France

### INFO ARTICLE

*Historique de l'article :*

Reçu le 17 février 2011

Accepté le 30 juillet 2012

Disponible sur Internet le 28 août 2012

Présenté par le Comité de rédaction

### RÉSUMÉ

Le problème d'intérêt est d'estimer la fonction de concentration et l'aire sous la courbe (AUC) à travers l'estimation des paramètres d'un modèle de régression linéaire avec un processus d'erreur autocorrélé. On introduit un estimateur linéaire sans biais simple et non paramétrique de la courbe de concentration et de l'AUC. On montre que cet estimateur construit à partir d'un plan d'échantillonnage régulier approprié est asymptotiquement optimal.

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

### ABSTRACT

The problem of interest is to estimate the concentration curve and the area under the curve (AUC) by estimating the parameters of a linear regression model with autocorrelated error process. We introduce a simple linear nonparametric unbiased estimator of the concentration curve and the AUC. We show that this estimator constructed from an appropriate regular sampling design is asymptotically optimal.

© 2012 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

## 1. Introduction

Dans beaucoup de situations expérimentales, telles que les études de la validité biologique, et en particulier dans les problèmes de la pharmacocinétique, où on étudie l'action des médicaments et leur emploi thérapeutique, on a souvent recours au calcul d'aire à partir d'un ensemble d'observations de mesures. Au cours de l'admission d'une dose à un agent pharmacologique, on estime l'aire sous la courbe (concentration–temps) connue sous le nom AUC, au vue de l'observation d'un échantillon de mesures accompagnées d'erreurs aléatoires. Pour une fonction de concentration, qui peut changer selon que le médicament est administré par voie orale ou intraveineuse, on cherche à déterminer un plan d'échantillonnage optimal pour un intervalle de temps fixé. L'aire engendrée par cet échantillonnage doit approcher au mieux l'aire sous la courbe de concentration (intégrale de la fonction de concentration sur l'intervalle de temps) qui est proportionnel à la biodisponibilité du médicament administré.

Plusieurs approches ont été proposées pour trouver les points d'échantillonnage optimaux du temps. Atkinson et al. [1] ont proposé une méthode d'optimisation en utilisant un critère de D-optimalité pour des paramètres comme l'aire sous la courbe de concentration (AUC), la concentration maximale du médicament et le temps associé. Katz et D'Argenio [7]

Adresses e-mail : [Mohamad.Belouni@imag.fr](mailto:Mohamad.Belouni@imag.fr) (M. Belouni), [Karim.Benhenni@upmf-grenoble.fr](mailto:Karim.Benhenni@upmf-grenoble.fr) (K. Benhenni).

ont proposé un algorithme qui utilise l'erreur quadratique moyenne (EQM) de l'estimation de l'AUC comme une fonction objective pour trouver les temps d'échantillonnage optimaux. Le but est de minimiser cette fonction objective par rapport aux temps d'échantillonnage afin de sélectionner le plan d'échantillonnage optimal. Kong et Gonin [8] et Wang [11] ont aussi utilisé le critère EQM pour estimer la courbe de concentration qui dépend du temps à travers une interpolation linéaire et déduire ainsi l'estimation de l'AUC. L'algorithme Recuit Simulé (SA) a été appliqué avec parcimonie dans ce domaine avec quelques exceptions notables. Par exemple, Jones et Wang [6], Duffull et al. [5] ont introduit l'algorithme (SA) pour trouver des points d'échantillonnage du temps avec des modèles utilisés en pharmacocinétique et également un critère de D-optimalité. En outre, ils ont comparé l'algorithme (SA) avec plusieurs autres algorithmes d'optimisation.

La règle des trapèzes est une règle de quadrature numérique qui est l'estimation non paramétrique la plus commune de l'aire sous la courbe de la concentration d'un médicament qui dépend du temps. Ce choix est très bien fondé, car cette règle donne en général une bonne approximation de l'AUC. Bailer et Piegorsch [2] ont comparé plusieurs méthodes de quadrature numérique pour approcher l'AUC en utilisant le critère EQM comme une fonction objective. Ils ont constaté que la règle des trapèzes a engendré un écart relativement plus petit entre l'AUC et son estimation.

Tous les travaux mentionnés dans cette introduction supposent que les observations sont indépendantes et donc que les erreurs sont non corrélées. La corrélation des erreurs peut avoir des conséquences importantes sur les propriétés statistiques des estimateurs des paramètres d'un modèle pour la fonction de concentration et de l'AUC, ainsi que sur le choix optimal du plan d'expérience.

Le problème d'intérêt est d'estimer la fonction de concentration  $g(t)$  et l'AUC à travers l'estimation des paramètres d'un modèle de régression linéaire. L'estimateur est construit à partir d'observations en un nombre fini de points d'un plan d'échantillonnage de manière appropriée lorsque le processus d'erreur prend des valeurs corrélées. La performance d'un estimateur est mesurée par son erreur quadratique moyenne (EQM). Les points d'échantillonnage qui minimisent l'EQM forment un plan d'échantillonnage optimal, qui est généralement difficile à engendrer. Pour contourner cette difficulté, Sacks et Ylvisaker [9] ont proposé une approche asymptotique et ont montré que les plans d'échantillonnage réguliers, qui sont déterminés par un bon choix de la densité d'échantillonnage, sont asymptotiquement optimaux lorsque le meilleur estimateur linéaire sans biais (BLUE) est utilisé et lorsque le processus d'erreur n'est pas régulier. Cependant, l'estimateur BLUE dépend de l'inverse de la matrice de covariance et par conséquent il est paramétrique et instable. Su et Cambanis [10] ont défini un estimateur simple construit à partir d'un plan d'échantillonnage médian modifié mais complexe. L'estimateur simple obtenu est biaisé mais asymptotiquement optimal. On introduit alors un estimateur linéaire sans biais simple et non paramétrique de  $g(t)$  et de l'AUC construit à partir d'un plan d'échantillonnage régulier non modifié. On montre que sous les mêmes hypothèses sur le processus d'erreur, cet estimateur linéaire sans biais construit à partir d'un plan d'échantillonnage régulier approprié est asymptotiquement optimal.

## 2. Modèle paramétrique et hypothèses

Pour une fonction de concentration  $g$  défini dans un intervalle d'observation  $[t_0, t_n]$ , on considère le modèle :

$$X(t_k) = g(t_k) + \varepsilon(t_k), \quad k = 0, 1, \dots, n,$$

où  $\{t_0, t_1, \dots, t_n\}$  sont les instants d'observation de  $X$  dans l'intervalle  $[t_0, t_n]$  et les erreurs  $\varepsilon(t_i)$  forment un processus aléatoire autocorrélé avec une moyenne nulle  $\mathbb{E}(\varepsilon(t)) = 0$  et une fonction d'autocovariance  $\mathbb{E}(\varepsilon(t)\varepsilon(s)) = R(t, s)$ . On fixe les points  $t_0$  et  $t_n$  et on s'intéresse au choix optimal des points intermédiaires  $\{t_1, \dots, t_{n-1}\}$  pour l'estimation de l'intégrale (AUC) suivante :

$$I(g) = \int_{t_0}^{t_n} g(t) dt.$$

On définit l'estimateur linéaire de  $I(g)$  par  $L_n(g)$ . Le but est de sélectionner les temps d'observations  $\{t_1^*, \dots, t_{n-1}^*\}$  qui minimisent l'erreur quadratique moyenne

$$\text{EQM}(L_n(g)) = \mathbb{E}(L_n(g) - I(g))^2.$$

On considère le modèle de régression linéaire multiple avec  $g(t) = \beta' f(t)$  où  $\beta' = (\beta_1, \dots, \beta_q)$  est un vecteur de paramètres inconnus et  $f'(t) = (f_1(t), \dots, f_q(t))$  est un vecteur de fonctions de régression connues.

On définit par  $h(t)$  une densité positive sur  $[0, 1]$  avec sa fonction de répartition  $H(t)$  et le plan d'échantillonnage régulier défini par :

$$\mathbf{T}_n = \left\{ t_{n,k} = H^{-1}\left(\frac{k}{n}\right), k = 0, \dots, n \right\}$$

qui comprend les bornes de l'intervalle  $[0, 1]$ ,  $t_{n,0} = 0$  et  $t_{n,n} = 1$ . En particulier si  $h \equiv 1_{[0,1]}$ , le plan d'échantillonnage régulier est périodique. Les plans d'échantillonnage réguliers ont été utilisés par exemple par Benhenni et Cambanis [3] pour l'estimation d'intégrale de processus stochastique et par Cambanis [4] pour les problèmes de détection du signal et de la régression simple.

Afin d'énoncer les résultats, on considère les hypothèses principales suivantes :

- (H1) Le processus d'erreur  $\varepsilon(t)$  est supposé n'avoir aucune dérivée en moyenne quadratique mais que sa fonction d'autocovariance  $R(t, s)$  est supposée avoir des dérivées partielles mixtes continues jusqu'à l'ordre deux en dehors de la diagonale ( $t \neq s$ ) du carré unité. Sur la diagonale ( $t = s$ ), on suppose que les limites à gauche et à droite  $R^{0,1}(t, t-) = \lim_{s \downarrow t} \partial R(t, s) / \partial s$ ,  $R^{0,1}(t, t+) = \lim_{s \uparrow t} \partial R(t, s) / \partial s$  existent et sont continues et la fonction de sauts  $\alpha(t) = R^{0,1}(t, t-) - R^{0,1}(t, t+)$  est supposée être strictement positive.
- (H2) Chaque fonction de régression  $f_j(t)$ ,  $j = 1, \dots, q$ , peut s'écrire sous la forme :  $f_j(t) = \int_0^1 R(t, s) \varphi_j(s) ds$  où  $\varphi_j$ ,  $j = 1, \dots, q$ , sont des fonctions continues connues sur  $[0, 1]$ .

On veut estimer la surface sous la courbe (AUC) de la fonction de concentration définie par

$$AUC = I(g) = \int_0^1 g(t) dt.$$

L'estimateur BLUE  $\hat{\beta}$  de  $\beta$  à partir d'observations continues du processus  $X$  sur tout l'intervalle  $[0, 1]$  est défini par

$$\hat{\beta} = S^{-1} \int_0^1 X(t) \varphi(t) dt,$$

où  $\varphi'(t) = (\varphi_1(t), \dots, \varphi_q(t))$ , et la matrice  $S$  de termes généraux  $s_{ij} = \int_0^1 \int_0^1 \varphi_i(t) R(t, s) \varphi_j(s) dt ds$ ,  $i, j = 1, \dots, q$ , est supposée être inversible. On estime alors l'AUC ( $I(g)$ ) par :  $L(g) = z' \hat{\beta}$  où  $z' = (z_1, \dots, z_q)$ ;  $z_j = \int_0^1 f_j(t) dt$ ,  $j = 1, \dots, q$ .

Si  $X(t)$  est observé en  $n + 1$  points d'un plan d'échantillonnage régulier  $T_n = \{t_{n,k}\}_0^n$  dans l'intervalle  $[0, 1]$ , alors on introduit l'estimateur non paramétrique sans biais du paramètre  $\beta$  par :

$$\hat{\beta}_n = S_n^{*-1} L_{n,q}(X),$$

où

$$S_n^* = (s_{ij}^*)_{q \times q}$$

est supposée être inversible et de termes généraux  $s_{ij}^* = \frac{1}{2n} \sum_{k=0}^{n-1} ((\frac{\varphi_i}{h} f_j)(t_{n,k}) + (\frac{\varphi_i}{h} f_j)(t_{n,k+1}))$ ,  $i, j = 1, \dots, q$ , et

$$L_{n,q}(X) = \frac{1}{2n} \sum_{k=0}^{n-1} \left( \left( \frac{\varphi X}{h} \right) (t_{n,k}) + \left( \frac{\varphi X}{h} \right) (t_{n,k+1}) \right).$$

Alors l'estimateur non paramétrique sans biais de l'AUC ( $I(g)$ ) est de la forme  $L_n(g) = z' \hat{\beta}_n = z' S_n^{*-1} L_{n,q}(X)$ .

**Théorème 2.1.** Si les hypothèses (H1) et (H2) sont vérifiées et si on suppose que  $\varphi_i/h$ ,  $i = 1, \dots, q$ , est deux fois continûment différentiable, alors l'estimateur  $\hat{\beta}_n$ , vérifie

$$\lim_{n \rightarrow \infty} n^2 (\text{EQM}(\hat{\beta}_n) - \text{trace}(S^{-1})) = \frac{1}{12} \int_0^1 \alpha(t) \frac{\varphi'(t) S^{-2} \varphi(t)}{h^2(t)} dt.$$

Par ailleurs, l'estimateur  $\hat{\beta}_n$  avec un plan d'échantillonnage  $T_n^*$  engendré par la densité

$$h^*(t) = \left\{ \alpha(t) \varphi'(t) S^{-2} \varphi(t) \right\}^{1/3} / \int_0^1 \left\{ \alpha(u) \varphi'(u) S^{-2} \varphi(u) \right\}^{1/3} du,$$

est asymptotiquement optimal.

**Théorème 2.2.** Sous les hypothèses (H1) et (H2) et si on suppose que  $\varphi_i/h$ ,  $i = 1, \dots, q$ , est deux fois continûment différentiable, alors l'estimateur  $L_n(g)$  de l'AUC, vérifie

$$\lim_{n \rightarrow \infty} n^2 (\text{EQM}(L_n(g)) - \text{EQM}(L(g))) = \frac{1}{12} \int_0^1 \alpha(t) \frac{\varphi'(t) (S^{-1} A S^{-1}) \varphi(t)}{h^2(t)} dt,$$

où  $A = z z'$ .

Par ailleurs, l'estimateur  $L_n(g)$  avec un plan d'échantillonnage  $\mathbf{T}_n^*$  engendré par la densité

$$h^*(t) = \{\alpha(t)\varphi'(t)(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1})\varphi(t)\}^{1/3} / \int_0^1 \{\alpha(u)\varphi'(u)(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}^{-1})\varphi(u)\}^{1/3} du,$$

est asymptotiquement optimal.

Dans le cas d'un modèle de régression linéaire simple  $X(t) = \beta f(t) + \varepsilon(t)$ , on obtient le résultat suivant.

**Corollaire 2.1.** Si  $q = 1$ , sous les mêmes hypothèses que Théorème 2.2, on obtient

$$\lim_{n \rightarrow \infty} n^2 (\text{EQM}(L_n(g)) - \text{EQM}(L(g))) = \frac{C}{12} \int_0^1 \alpha(t) \frac{\varphi^2(t)}{h^2(t)} dt,$$

où

$$C = \frac{(\int_0^1 f(t) dt)^2}{(\int_0^1 f(t)\varphi(t) dt)^2}$$

et le plan d'échantillonnage  $\mathbf{T}_n^*$  engendré par la densité

$$h^*(t) = \{\alpha(t)\varphi^2(t)\}^{1/3} / \int_0^1 \{\alpha(u)\varphi^2(u)\}^{1/3} du,$$

est asymptotiquement optimal.

## Références

- [1] A.C. Atkinson, K. Chaloner, A.M. Herzberg, J. Juritz, Optimum experimental designs for properties of a compartmental model, *Biometrics* 49 (1993) 325–337.
- [2] A.J. Bailer, W.W. Piegorisch, Estimating integrals using quadrature methods with an application in pharmacokinetics, *Biometrics* 46 (4) (1990) 1201–1211.
- [3] K. Benhenni, S. Cambanis, Sampling designs for estimating integrals of stochastic processes, *Annals of Statistics* 20 (1992) 161–194.
- [4] S. Cambanis, Sampling designs for time series, in: E.J. Hannan, P.R. Krishnaiah, M.M. Rao (Eds.), *Handbook of Statistics*, in: Time Series in Time Domain, vol. 5, North-Holland, Amsterdam, 1985, pp. 337–362.
- [5] S.B. Duffull, S. Retout, F. Mentré, The use of simulated annealing for finding optimal population designs, *Computer Methods and Programs in Biomedicine* 69 (2002) 25–35.
- [6] B. Jones, J. Wang, Constructing optimal designs for fitting pharmacokinetic models, *Statistics and Computing* 9 (1999) 209–218.
- [7] D. Katz, D.Z. D'Argenio, Experimental design for estimating integrals by numerical quadrature with applications to pharmacokinetic systems, *Biometrics* 39 (1983) 621–628.
- [8] F.H. Kong, R. Gonin, Optimal sampling times in bioequivalence tests, *Journal of Biopharmaceutical Statistics* 10 (2000) 31–44.
- [9] J. Sacks, D. Ylvisaker, Designs for regression problems with correlated errors, *Annals of Mathematics* 37 (1966) 66–89.
- [10] Y. Su, S. Cambanis, Sampling designs for regression coefficient estimation with correlated errors, *Annals of the Institute of Statistical Mathematics* 46 (1993) 707–722.
- [11] J. Wang, Optimal design for linear interpolation of curves, *Statistics in Medicine* 20 (2001) 2467–2477.