NONPARAMETRIC APPROACH USING GAUSSIAN KERNELS ESTIMATES JOINT PROBABILITY DENSITIES IN POPULATION PHARMACOKINETICS

H. MAROUANI [1], L. CLARET [2], C. FAIVRE [1], A. ILIADIS [1]
[1] Dpt. of Pharmacokinetics, UMR-MD3, Univ. Méditerranée, Marseille, France
[2] Pharsight Corp., Mountain View, USA

Introduction

- Single- and two-stage methods intended to process data for PK/PD population studies. Single-stage methods (nonlinear mixed effect models) process on raw data to estimate population characteristics. Two-stage methods performs PK/PK modeling in the first stage leading to the training data and these are processed within the statistical model in the second stage.
- Both methods are associated to parametric or nonparametric approaches for the statistical description of data.
- Two-stage methods require extensive sampling and fail to combine correctly the variabilities underlying the data.

Objective

- Describe the interindividual variability by means of a multivariate probability density function (pdf) using a two-stage nonparametric procedure.
- Our proposal (1) enables incorporating the residual variability (including intra-individual variability and measurements errors) and out-performs the traditional two-stage method, which ignores residual variability and overestimates random effects; (2) improves performances of the multivariate ordinal kernel density estimation; (3) selects the optimal bandwidth in the sense of the minimum approximated mean integrated squared error.

Materials and Methods

Individual kinetic parameters

- We assume the error between observed concentrations y_j at times t_j and the predicted values by the theoretical model $y_M(t_j, x)$ follows an additive normal distribution with an heteroscedastic error variance model. The model involves p parameters denoted x.

- If the errors are independent at two different times, then the likelihood function evaluated over the m sampling points can be written as:

$$L(y/x) = (2\pi)^{-m/2}|V|^{-1/2}\exp\left[-(y-y_M)^T V^{-1}(y-y_M)/2\right] \quad (1)$$

y and y_M are m-dimensional vectors with components y_j and $y_M(t_j, x)$, respectively. V is the m-order diagonal matrix of error variances.

- The maximum likelihood estimate (MLE) $\hat{x}$ is the value of x that gives the particular observation y the higher probability of occurrence. MLE is consistent and asymptotically efficient and Gaussian, i.e., $\hat{x}$ is distributed normally around the true value x_0 as:

$$g(\hat{x}) = (2\pi)^{-p/2}|P|^{-1/2}\exp\left[-(\hat{x}-x_0)^T P^{-1}(\hat{x}-x_0)/2\right] \quad (2)$$

where P is the covariance matrix of estimates, given by the inverse of the expected information matrix, and provides a measure of estimation precision.

- Over the n available individuals, MLE leads to the training data that are $\hat{x}_i$ estimates and P_i precisions.

Selecting optimal bandwidth

- The choice of the bandwidth s in the relationship (4) is critical in terms of performance when implementing kernel approaches.

- The mean integrated squared error (MISE) measures the distance between the estimated $\hat{f}(x;s)$ and the unknown density $f(x)$. The optimal bandwidth $\hat{s}$ minimizes MISE. Bandwidth selectors proceed on an approximated MISE obtained under asymptotic conditions [1,2].

- The implemented selector is the least-squares cross-validation procedure [1]. The basic principle is to construct the approximated MISE from the data themselves (Jackniffe) and evaluate it over a grid of s values.

Simulation study

- One compartment model with $p=2$ parameters, volume of distribution and clearance. Reference interindividual distribution: mixture of multivariate normal and log-normal distributions:

$$f(x) = \alpha \cdot N(\mu_N, \Sigma_N) + LG(\mu_{LN}, \Sigma_{LN})$$

with

$$\mu_N = [60\ 15]^T \qquad \Sigma_N = \begin{bmatrix} 15 & 1.71 \\ 1.71 & 4 \end{bmatrix}$$

$$\mu_{LN} = [3.4\ 0.3]^T \qquad \Sigma_{LN} = \begin{bmatrix} 0.03 & -6.64E-2 \\ -6.64E-2 & 0.3 \end{bmatrix}$$

$$\alpha = 2/3 \qquad \rho_N = 0.7 \qquad \rho_{LN} = -0.7$$

- $N = 500$ individuals were drawn. Each individual randomly received 2, 5, 7, 10, or 20 mg of drug given by infusion during 0.5, 1, 2, or 3 h (randomly associated). Three different sampling protocols of size $m = 7$ were randomly associated to individuals. Simulated concentrations were disturbed by heteroscedastic measurement errors with CV of 5, 10, or 15% (randomly associated to each kinetic profile).

- From the disturbed kinetic profiles, PK parameters were obtained using MLE and a quasi-Newton optimization routine. The training data are $\hat{x}_i$ P_i $i = 1:N$.

Kernel Density Estimation

- The ordinary kernel estimation of prior pdf $f(x)$ is given by:

$$\hat{f}(x;h) = n^{-1}\sum_{i=1}^{n} K_h(x - x_i) \quad (3)$$

The kernel function K_h with bandwidth h is defined by $K_h(z) = h^{-p}K(z/h)$.

- The basic idea of the present study is to use the training data [3] and:
  1. center kernels at the individual estimates $\hat{x}_i$,
  2. allow the bandwidth of the kernels to vary from one individual to another according to P_i which is associated with the reliability to each estimate.

- Because of (2), the Gaussian kernel is used. Therefore, the bandwidth h should depend on the individual precision of estimates P_i, i.e., $h_i = sP_i^{1/2}$. s is the new bandwidth.

- To allow smoothing equally scaled in all directions by a single s, the data were "pre-whiten" by the covariance matrix D which expresses the dispersion of individuals weighed by P_i. This involves the following:
  1. Obtain $\tilde{P}_i$ the Choleski factor of P_i, i.e., $P_i = \tilde{P}_i^T \tilde{P}_i$.
  2. Transform x_i to z_i by $z_i = \left[\tilde{P}_i^{-1}\right]^T x_i$.
 3. Compute the covariance matrix D of z_i.

- The final form of (3) is:

$$\hat{f}(x;s) = n^{-1}(2\pi)^{-p/2}s^{-p}\sum_{i=1}^{n}|A_i|^{-1/2}\exp\left[-\frac{1}{2s^2}(x-x_i)^T A_i^{-1}(x-x_i)\right] \quad (4)$$

with $A_i = \tilde{P}_i^T D \tilde{P}_i$. The matrices A_i embody individual precision (originated from intra-individual variability and measurement errors), and gives the same importance in all directions of the parameter space.

Results

- Variable number of individuals in the data base was considered: $n = 50:10:N$.
- Grid of 200 s values in the range 0.1:0.4 for cross validation.
- 350 individuals describe reliably the data (subplot 1).
- Optimal bandwidth $\hat{s} = 0.120$ for $n = 350$ (subplot 2).
- 3D plot of the reference distribution $f(x)$ (subplot 3).
- 3D plot of the estimated distribution $\hat{f}(x;\hat{s})$ (subplot 4).

Reference distribution

Estimated distribution
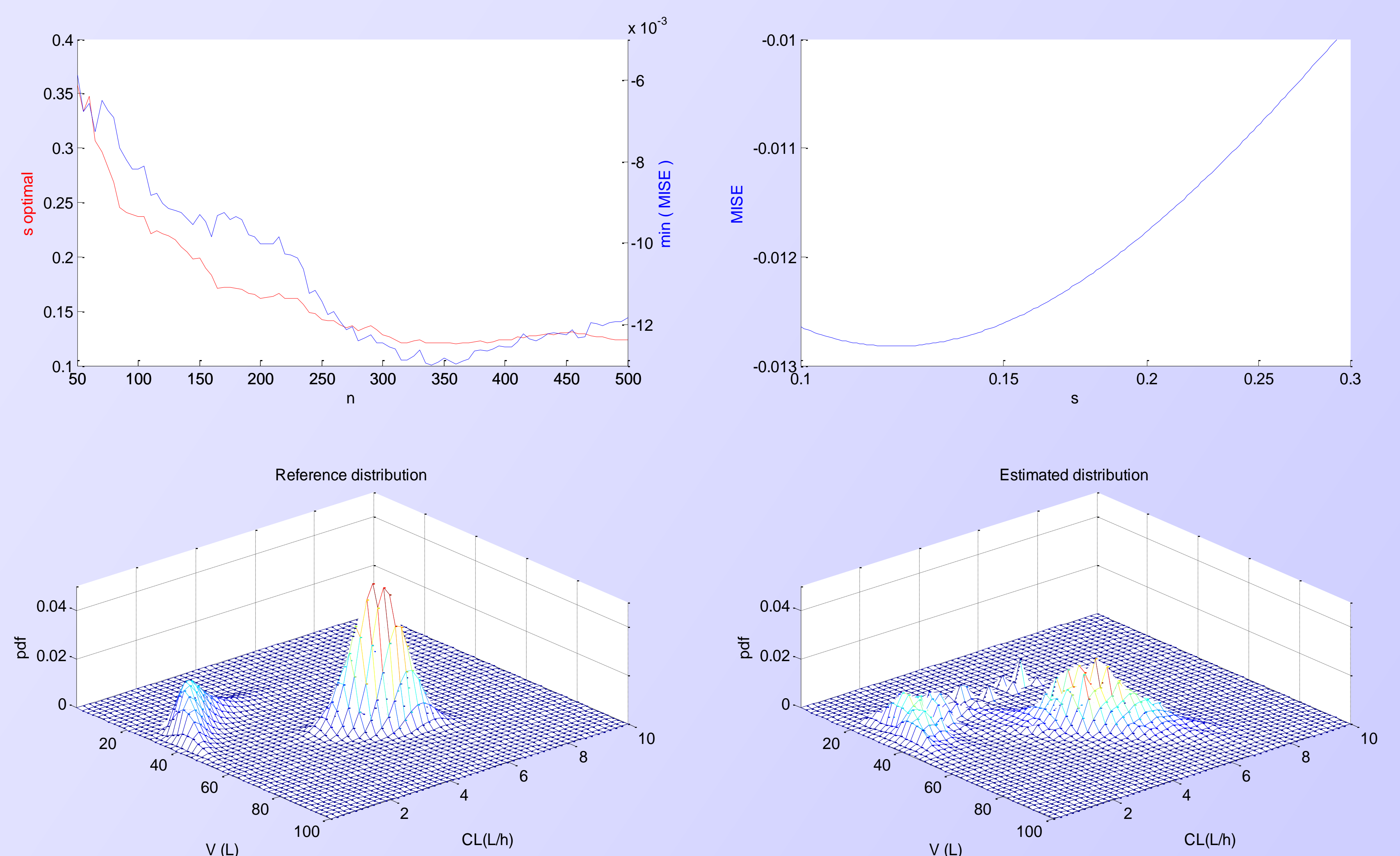
Conclusion

- Both residual and interindividual variability are incorporated in the proposed nonparametric approach. The optimal bandwidth in the kernel pdf estimator is determined. From a practical point of view, the method is easy to implement and quick for data processing. In order to achieve validation, the performances of this new estimator must be compared to other kernel procedures.

- Since contaminated data are considered, one should also develop the deconvolving kernel estimator [4]. This estimator need very challenging theoretical analysis and the practical computation is time consuming: it is under investigation.

References

1. Silverman BW. Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
2. Wand MP, Jones MC. Kernel smoothing, Chapman and Hall, London, 1995.
3. Claret L, Iliadis A. Non parametric Density estimation applied to population Pharmacokinetics. Math Biosci. 1996 Apr 1;133(1):51-68.
4. Hazelton Martin L, Turlach Berwin A. Nonparametric density deconvolution by weighted kernel estimators. Stat Comput 2009; 19: 217–228