CONFIDENCE BANDS FOR REGRESSION: THE INDEPENDENCE POINT METHOD

A. J. HAYTER, H. P. WYNN, W. LIU

ABSTRACT

In some circumstances for a linear model with □ parameters □ regression in Rd of the form □ one can find special points □ for which the usual least squares estimators □ of the expected response □ are uncorrelated, independent in the Gaussian case. Following Wynn (Biometrika, 1984) we use this to set up simple piecewise linear confidence bands in the case □, namely the additive main effect model in multiple regression and some other cases.

(this abstract contains LaTex markup that cannot be reproduced here; see the original paper for details)

Southampton Statistical Sciences Research Institute
Methodology Working Paper M05/17

Confidence bands for regression: the independence point method

A. J. Hayter
Industrial and Systems Engineering, Georgia Tech, Atlanta, USA
ajh@isye.gatech.edu

H. P. Wynn
Department of Statistics, London School of Economics, London, UK
h.wynn@lse.ac.uk

W. Liu
School of Mathematics, University of Southampton, Southampton, UK
w.liu@maths.soton.ac.uk

October 2004

Summary

In some circumstances for a linear model with k parameters $\theta = (\theta_0, \ldots, \theta_{k-1})$ regression in R^d of the form

$$Y_x = f(x)^T \theta + \epsilon_x$$

one can find special points $z^{(1)}, \ldots, z^{(k)}$ for which the usual least squares estimators $\hat{\eta}(z^{(r)})$ of the expected response $\eta(z^{(r)}) = f(z^{(r)})^T \theta$, $(r = 1, \ldots, k)$ are uncorrelated, independent in the Gaussian case. Following Wynn (Biometrika, 1984) we use this to set up simple piecewise linear confidence bands in the case $f(x) = (1, x_1, \ldots, x_{k-1})$, namely the additive main effect model in multiple regression and some other cases.

Key words: linear regression; confidence bands; moment theory; gaussian quadrature

1 Introduction

There is a considerable literature on confidence bands in linear regression. The starting point is the standard elliptically shaped confidence region band for a set of model parameters, Scheffé (1953, 1959) and the realisation that it is by no means easy to go from this to simultaneous band for a regression *function*. The problem is that if the band is required to hold over restrict function class or over a restricted region then confidence regions induced in parameter space are often complex. This means that the width required to yield an exact pre-specified coverage probability is hard to compute. An early paper by Wynn and Bloomfield (1971) addresses the problem for straight line (simple) linear regression over an interval and one-dimensional quadratic regression over the whole real line and draws on papers by Gafarian (1964) and Graybill and Bowden (1967). A recent paper on the one-sided confidence bands is Pan et al (2003). When a regression model has more than one explanatory variable, exact simultaneous confidence bands are available for only few special situations, e.g. Casella and Strawderman (1980) constructed an exact confidence band over a region defined by a second order inequality of the explanatory variables. Most published work is on either conservative confidence bands, e.g. Naiman (186,1990) developed tube theory approximation based on early work by Working and Hotelling (1929) and Weyl (1939), or approximate confidence bands, e.g. Sun, Loader and McCormick (2000) used the approximate distribution of the Gaussian maxima. A recent paper on the constrained region problem is Liu et al (2004), which uses Monte Carlo simulation to produce pseudo exact simultaneous confidence bands. Some general references to simultaneous inference and the the almost synonymous multiple comparisons are Tamhane and Hochberg (1987) and Hsu (1996).

The purpose of this paper is to provide a method that allows easy construction of exact simultaneous confidence bands, especially for the additive main effect model in multiple linear regression. The proposed method is a generalisation to R^d of the one dimensional method of Wynn (1984). We write a linear regression as:

$$Y_x = f(x)^T\theta + \epsilon_x,$$

where Y_x is the response at a point x in R^d and $f(x) = (f_0(x), \dots, f_{k-1}(x))^T$ is a vector of, typically continuous functions, and ϵ is an errors and the ϵ_i are uncorrelated with equal variance σ^2. In an actually experiment, or observational study, we observe $Y_{x^{(i)}} = Y_i$ at points $D = \{x^{(i)},\ i = 1, \dots, n\}$ and write

$$Y = X\theta + \epsilon$$

where Y is the vector of observations, $X = \{f_j(x^{(i)})\}$ and ϵ is the vector of errors. Under standard assumptions the least squares estimator of θ is $\hat{\theta} = (X^TX)^{-1}X^TY$ and the predictor of the expected response $\eta(x) = E(Y_x) = f(x)^T\theta$ is $\hat{\eta}(x) = f(x)^T\hat{\theta}$.

The method is based on the fact that for some design model pairs $\{D, f(x)\}$, one can find a set of points $z^{(r)} \in R^d$ such that the corresponding $\hat{\eta}(z^{(r)})$ are independent. We will then use this in the construction of confidence bands.

Definition 1 *For a standard regression set-up with model function $f(x)$ points $z^r \in F^d$, $(r = 1, \ldots, m)$, such that the standard least squares predictors $\hat{\eta}(z^{(r)})$, $(r = 1, \ldots, m)$ are uncorrelated (independent in the Gaussian case) are called m* independence points.

1.1 Construction of confidence bands

We are interested in constructing simultaneous bands $B(x) = (\underline{b}(x, Y), \bar{b}(x, Y))$ for $\eta(x) = f(x)^T\theta$ with simultaneous coverage probability $1 - \alpha$ for x in some region $R \subset R^d$:

$$\text{prob}_Y\{\underline{b}(x, Y) \leq \eta(x) \leq \bar{b}(x, Y), \text{ for all } x \in R \mid \theta\} = 1 - \alpha. \tag{1}$$

It is important to note that the coverage probability depends on the set R. For example for two sets R_1 and R_2 with $R_1 \subset R_2$ and a fixed band the coverage probability is not less for R_1, because the corresponding restrictions are not more. Note that we suppress Y in the $b(x, Y)$ notation.

A band for a given set R_1 can be extended to a set R_2 using the following construction. Define the set

$$\Theta = \{\theta | \underline{b}(x) \leq f(x)^T\theta \leq \bar{b}(x), \text{ for all } x \in R_1\}. \tag{2}$$

It is clear that Θ is a convex set, because for $0 \leq \beta \leq 1$

$$\eta = (1 - \beta)f(x)^T\theta_1 + \beta f(x)^T\theta_2 = f(x)^T\{(1 - \beta)\theta_1 + \beta\theta_2\}$$

so that if $\eta_1 = f(x)^T\theta_1$ and $\eta_2 = f(x)^T\theta_2$ both satisfy the band so does $\eta(x)$.

It may be that the band in (1) have redundancy, that is to say narrower bands give the same Θ. This leads to the following definition originating in Wynn and Bloomfield (1971).

Definition 2 *For a set $R = R_1$, a band (as described in (1)) is said to be* taut *if*

$$\underline{b}(x) = inf_{\theta \in \Theta_1}\eta(x), \quad \bar{b}(x) = sup_{\theta \in \Theta_1}\eta(x)$$

where

$$\Theta_1 = \{\theta | \underline{b}(x) \leq \eta(x) = f(x)^T\theta \leq \bar{b}(x), \textit{ for all } x \in R_1\}.$$

Now suppose the band in (1) is taut and let $R_2 \supset R_1$. Construct the band for R_2 using the upper and lower envelopes under Θ_1, extended to R_2:

$$\underline{b}_2(x) = \inf_{\theta \in \Theta_1} f(x)^T\theta, \quad \bar{b}_2(x) = \sup_{\theta \in \Theta_1} f(x)^T\theta, \text{ for all } x \in R_2$$

and

$$\underline{b}_2(x) \leq \eta(x) \leq \bar{b}_2(x), \text{ for all } x \in R_2 \tag{3}$$

The bands clearly agree on R_1, by construction, and are determined by the same set Θ_1. The coverage probability is also the same: $1-\alpha$.

This paper bases the bands on independence points, according to Definition 1 . Let $R_1 = \{z_1, \ldots, z_k\}$ be a set of independence points. Construct the band for R_1 by taking

$$\underline{b}(x) = \hat{\eta}(x) - c(x)s, \;\; \bar{b}(x) = \hat{\eta}(x) + c(x)s, \;\; x \in R_1 \tag{4}$$

where $c(x)$ is determined partly to set the coverage at $1-\alpha$ and partly to control the shape of the bound and s, is a quantity possibly dependent on Y, but not on x. We need to prove that the band is taut. But this must be case, since, from the Definition 1, $f(z_r), (r = 1, \ldots, k)$ are algebraically independent as vectors so that the *inf* and *sup* in Definition 2 can be achieved, from simple geometric considerations.

The simplicity of the construction means that the coverage probability is straightforward to evaluate. We require under the usual Gaussian assumption for the statement (4) to have probability $1-\alpha$. Suppose $\mathrm{var}(\hat{\eta}(x)) = \sigma^2 v(x)$. Then one construction is to take s to be the usual unbiased independent estimate of σ, $c(x) = c_{1-\alpha}\sqrt{v(x)}$ and $c_{1-\alpha}$ chosen to obtain the correct coverage probability, $1-\alpha$. The independence property of the $\hat{\eta}(x), \;\; x \in R_1$ means that $c_{1-\alpha}$ is simply the critical value for the maximum modulus statistic, which involves at most two dimensional numerical integrations and has been well tabulated (e.g. Hochberg and Tamhane, 1987).

1.2 The link with moment theory

Let $z_1, \ldots, z_k$ be independence points. We write down the condition which arises from Definition 1. This is simply that the covariance matrix of the $\hat{\eta} = (\hat{\eta}(z_1), \ldots, \eta(z^{(k)})^T$ is:

$$\mathrm{cov}(\hat{\eta}) = \sigma^2 Z(X^T X)^{-1} Z^T = \sigma^2 K, \tag{5}$$

where $Z = (f(z_1), \cdots, f(z_k))^T$, K is diagonal and σ^2 is the error variance.

Inverting this equation and dividing by the sample size we obtain

$$\frac{1}{n} X^T X = \frac{1}{n} Z^T K^{-1} Z. \tag{6}$$

The left hand side of (6) is the (cross-product) moment matrix for the functions $f_j(x), (j = 0, \ldots k-1,)$ and uniform discrete measure at the design points, which we call the *design measure*. The right hand side is the moment matrix for the measure with on the support $D^* = \{z_1, \ldots, z_k\}$ with weights $k_r = \frac{1}{n} K_{rr}, \;(r = 1, \ldots, k)$. Thus, the existence of depth k independence points is equivalent to the solution of a moment problem: does there a measure on points $z_1, \ldots, z_k$ with the same moment matrix as for the original design? This problem is sometimes referred to as *moment matching*.

2 Additive main effect regression

Consider the case in which $d = k - 1$, and the regression function:

$$f(x)^T = (1, x_1, ..., x_{k-1}),$$

so that the regression is the simple main effect model

$$Y_x = \theta_0 + \theta_1 x_1 + \ldots + \theta_{k-1} x_k + \epsilon_x.$$

Let the design be $D = \{x^{(1)}, \ldots, x^{(n)}\}$ and let

$$U = (x^{(1)} : \ldots : x^{(n)})^T$$

be the $n \times (k-1)$ matrix holding the design points as rows. Let j_r be the r-vector of ones. Then the X-matrix is

$$X = [j_n : U]$$

and

$$X^T X = \begin{bmatrix} n & j_n^T U \\ U^T j_n & U^T U \end{bmatrix}.$$

2.1 Independence points

We want to find independence points $z^{(1)} \ldots z^{(k)}$ such that (6) holds. We shall proceed in stages. First we centre the model by subtracting the column means, taking care to express the operations in matrix terms. Thus define

$$\overline{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_j^{(i)}$$

and write the centred regression function as

$$\tilde{f}(x) = (1, x_1 - \overline{x}_1, \ldots, x_{k-1} - \overline{x}_{(k-1)})^T.$$

Then define the matrix

$$T = \begin{bmatrix} 1 & -\overline{x}_1 & \ldots & -\overline{x}_{(k-1)} \\ 0 & 1 & 0 & \ldots 0 \\ \ldots & \ldots & & \ldots \\ 0 & \ldots & 0 & 1 \end{bmatrix}.$$

The transformed model is $\eta(x) = \tilde{f}(x)^T \phi$ and in matrix terms we have

$$\eta = X\theta = \tilde{X}\phi$$

where

$$\tilde{X} = XT = [j_n : V],$$

and

$$V = \{x_j^{(i)} - \overline{x}_j\}.$$

We work with $\tilde{X}$ and recapture the result for X at the end. By construction

$$\frac{k}{n}\tilde{X}^T\tilde{X} = \begin{bmatrix} k & \underline{0}^T \\ \underline{0} & \frac{k}{n}V^TV \end{bmatrix}.$$

Since X^TX is full rank it follows that V^TV is also full rank. Write the Cholesky factorisation of $\frac{k}{n}V^TV$ as

$$\frac{k}{n}V^TV = C^TC, \tag{7}$$

where C is $(k-1)\times(k-1)$ (we can use any suitable factorisation). The completion of the proof relies on a simple fact: we can always find a $k \times (k-1)$ matrix H such that

$$j_k^T H = 0, \quad H^TH = I_{k-1},$$

where I_{k-1} is the $(k-1)\times(k-1)$ identity. In other words take the columns of H to be *any* $k-1$ orthonormal vectors which are also orthogonal to j_k. With such a H rewrite (7) as

$$\frac{k}{n}V^TV = C^TH^THC,$$

and note that since $j_k^TH = 0$ we have $j_k^THC = 0$. This means that if we write $W = HC$ and

$$\tilde{H} = [j_k : W]$$

we have

$$\tilde{H}^T\tilde{H} = \frac{k}{n}\tilde{X}^T\tilde{X}.$$

Now, returning to X

$$X^TX = (T^T)^{-1}\tilde{X}^T\tilde{X}T^{-1}.$$

But

$$T^{-1} = \begin{bmatrix} 1 & \overline{x}_1 & \dots & \overline{x}_{(k-1)} \\ 0 & 1 & 0 & \dots 0 \\ \dots & \dots & & \dots \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

and

$$\frac{k}{n}X^TX = (T^T)^{-1}\tilde{H}^T\tilde{H}T^{-1}$$

The first column of $\tilde{H}T^{-1}$ is j_k so write

$$\tilde{H}T^{-1} = [j_k : Q]$$

where $Q = WT^{-1} = HCT^{-1}$.

Hence

$$X^T X = \frac{n}{k}[j_k : Q]^T [j_k : Q],$$

which is exactly in the form required in (6) with the diagonal matrix K being simply diag$\{\frac{k}{n}\}$. The independence points are the rows of Q, and the variances of the $\hat{\eta}(z^{(r)})$ are $\frac{k}{n}\sigma^2$ since $K = \text{diag}\{\frac{k}{n}\}$.

We illustrate the construction above with the case $d = k - 1 = 2$ in detail. The model is

$$Y_{(x_1,x_2)} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \epsilon.$$

Starting with the centred case we have

$$\frac{3}{n}\tilde{X}^T\tilde{X} = 3\begin{bmatrix} 1 & 0 & 0 \\ 0 & s_{20} & s_{11} \\ 0 & s_{11} & s_{02} \end{bmatrix},$$

$$s_{20} = \frac{1}{n}\sum_i (x_1^{(i)} - \bar{x}_1)^2,\ s_{02} = \frac{1}{n}\sum_i (x_2^{(i)} - \bar{x}_2)^2,\ \ s_{11} = \frac{1}{n}\sum_i (x_1^{(i)} - \bar{x}_1)(x_2^{(i)} - \bar{x}_2)$$

A Cholesky factor, C, of

$$S = 3\begin{bmatrix} s_{20} & s_{11} \\ s_{11} & s_{02} \end{bmatrix} = C^T C$$

is

$$C = \sqrt{3}\begin{bmatrix} \sqrt{s_{20}} & \frac{s_{11}}{\sqrt{s_{20}}} \\ 0 & \sqrt{\frac{s_{20}s_{02} - s_{11}^2}{s_{20}}} \end{bmatrix}.$$

We take

$$H = \begin{bmatrix} 0 & \sqrt{\frac{2}{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \end{bmatrix},$$

giving

$$W = \sqrt{3}\begin{bmatrix} 0 & \sqrt{\frac{2}{3}}\sqrt{\frac{s_{20}s_{02}-s_{11}^2}{s_{20}}} \\ \sqrt{\frac{s_{20}}{2}} & \frac{1}{\sqrt{2}}\frac{s_{11}}{\sqrt{s_{20}}} - \frac{1}{\sqrt{6}}\sqrt{\frac{s_{20}s_{02}-s_{11}^2}{s_{20}}} \\ -\sqrt{\frac{s_{20}}{2}} & -\frac{1}{\sqrt{2}}\frac{s_{11}}{\sqrt{s_{20}}} - \frac{1}{\sqrt{6}}\sqrt{\frac{s_{20}s_{02}-s_{11}^2}{s_{20}}} \end{bmatrix};$$

It is easily confirmed that $W^T W = S$ and $(1,1,1)^T W = 0$. Transforming back to the original space we have the three independence points, which are the rows of W plus the mean vector $(\overline{x}_1, \overline{x}_2)$.

If we specialize further and take the design to be the 2^2 full factorial design $\{(\pm 1, \pm 2)\}$ the independence points are

$$(0, \frac{2}{\sqrt{3}}), \quad (1, -\frac{1}{\sqrt{3}}), \quad (-1, -\frac{1}{\sqrt{3}}). \tag{8}$$

These lie on a circle centred at the origin and with radius $\frac{2}{\sqrt{3}}$. This arises from a Fourier construction, which applies generally. Let $k = 2m + 1$ and take the vector

$$\frac{\sqrt{2}}{k}(\sin(2\pi x), \cos(2\pi x), \sin(4\pi x), \cos(4\pi x), \ldots \sin(2k\pi x), \cos(2k\pi x))$$

evaluated at the equally spaced points

$$x = 0, \frac{1}{k}, \frac{2}{k}, \ldots \frac{k-1}{k} \in [0, 1].$$

Then it is well known from elementary Fourier analysis that these vectors are orthonormal and are orthogonal to j_k and may therefore be taken as the rows of H. The case $k = 2m$ can be handled with a small adaptation for the last vector.

2.2 The band

Let $R_1 = \{z^{(1)}, \ldots, z^{(d+1)}\}$ be the independence points found in Section 2.1. Construct the band on R_1 by taking

$$\underline{b}(z^{(i)}) = \hat{\eta}(z^{(i)}) - c_i \mathrm{Var}(\hat{\eta}(z^{(i)})s, \ \ \bar{b}(z^{(i)}) = \hat{\eta}(z^{(i)}) + c_i \mathrm{Var}(\hat{\eta}(z^{(i)})s,$$

$i = 1, \cdots, d + 1$, where s is the usual unbiased estimator of σ, and the constants $c_1, \cdots, c_{d+1}$ are chosen such that

$$P\{\underline{b}(z^{(i)}) \leq \eta(z^{(i)}) \leq \bar{b}(z^{(i)}), \ z^{(i)} \in R_1\} = P\{|N_i| \leq c_i \sigma^{-1} s, \ i = 1, \cdots, d+1\} = 1-\alpha \tag{9}$$

where $N_1, \cdots, N_{d+1}$ are i.i.d. standard normal random variables independent of s. It is clear that the c_i are easily computed from (9). Especially when the c_i are set to be equal then their common value is simply given by the maximum modulus distribution.

Now that an exact simultaneous confidence band on R_1 is available, it can be extended through (3) to any region R_2 that contains R_1, especially to the whole space $R_2 = R^d$. Note that the points $z^{(1)}, \ldots, z^{(d+1)}$ are in general positions, that is not on a lower affine subspace, except that we restrict the vector means $(\overline{z}_1, \ldots, \overline{z}_{(d+1)})$ to be the origin . These form a simplex $\mathcal{S}$. Add a $(d + 1)$-th dimension, which we call the y-dimension. Define the following region in R^{d+1}

$$B = \{(z, z^T\theta) : \ \underline{b}(z^{(i)}) \leq {z^{(i)}}^T\theta \leq \bar{b}(z^{(i)}), \ i = 1, \ldots, d + 1, \ \theta \in R^{d+1}\}.$$

The region B has upper and lower boundaries which are describe as follows and form respectively the upper and lower surfaces the confidence band on R^d. On the simplex $\mathcal{S}$, the upper and lower boundaries are given respectively by the simplex $\mathcal{S}^+$ formed from the points $(z^{(i)}, \bar{b}(z^{(i)})), \ i = 1, \ldots, d + 1$ and the simplex $\mathcal{S}^-$ formed from the points $(z^{(i)}, \underline{b}(z^{(i)})), \ i = 1, \ldots, d + 1$. Outside $\mathcal{S}$ the upper boundary is formed by the supreme of the d-dimensional hyperplanes which pass through h points $(z^{(i)}, \bar{b}(z^{(i)})), \ i = 1, \ldots, h$ and $d + 1 - h$ points $(z^{(i)}, \underline{b}(z^{(i)})), \ i = h+1, \ldots, d+1$. The lower boundary is formed similarly. Figure 1 shows a schematic of the contours of the upper surface of the band based on the three points in (8), assuming $\hat{\eta}(z^{(i)}) = 0, \ i = 1, 2, 3$.

Figure 1: Independence point band contours: linear model

3 The quadrature method

Let us recall, briefly the quadrature method for solving (6), for one dimensional polynomial regression as give in Wynn (1987). Let $d = 1$, let $p_{2n+1}(x)$ be a polynomial of degree $2n + 1$ and q_{n+1} an orthogonal polynomial with respect to the discrete measure which puts uniform mass at the points of a design D. Then dividing p_{2n+1} by $q_{n+1}(x)$ we can write

$$p_{2n+1}(x) = s_n(x)q_{n+1}(x) + R(x), \tag{10}$$

where $s(x)$ and $R(x)$ have degree at most n. Integrating with respect to the discrete design measure and using the fact that q_{n+1} is orthogonal to any polynomial of lower degree we have

$$\frac{1}{n}\sum_{x\in D} p_{2n+1}(x) = \frac{1}{n}\sum_{x\in D} R(x). \tag{11}$$

Let $z^{(1)}, \ldots, z^{(n+1)}$ be the zeros of q_{n+1} and let $L_r(x)$ be the standard degree n indicator functions used in Lagrange interpolation on $z^{(1)}, \ldots, z^{(n+1)}$:

$$L_r(x) = \frac{\prod_{s=1,;\, s\neq r}^{n+1}(x - z^{(s)})}{\prod_{s=1,;\, s\neq r}^{n+1}(z^{(r)} - z^{(s)})}.$$

Then we can express $R(x)$ uniquely in terms of the L_r:

$$R(x) = \sum_{r=1}^{n} R(z^{(r)})L_r(x)$$

so that integrating the right hand side of we obtain the quadrature formula:

$$\frac{1}{n}\sum_{x\in D} p_{2n+1}(x) = \sum_{r=1}^{n+1} w_r R(z^{(r)})$$

where

$$w_r = \frac{1}{n}\sum_{x\in D} L_r(x).$$

If we apply this in the case where p_{2n+1} is all the monomial $x^k,\ k=1,\dots,2n$ we obtain (6).

We note several features of this analysis: the use of orthogonal polynomials, their zeros and the indicator basis for the remainder over the zeros. This can be generalised using some of the ideas from "algebraic statistics", see Pistone et al (2000). We present a simple version of the method, without all the algebraic background.

Give a design D in R^d we select a saturated monomial basis $\{x^\alpha,\ \alpha\in L\}$ where we have used the notation $x^\alpha = x_1^{\alpha_1}\cdots x_d^{\alpha_d}$ and $|D|=|L|=n$. This means that we can fit a model

$$\eta(x) = \sum_\alpha \theta_\alpha x^\alpha,$$

and the $n\times n$ X-matrix $\{x^\alpha\}_{x\in D,\,\alpha\in L}$ is nonsingular. The algebraic methods provide a way doing this selection, but we omit the details.

Select a total (linear) ordering of the elements of L (typically consistent with increasing degree) and let $f(x)$ be the $n\times 1$ vector of monomials $x^\alpha,\ \alpha\in L$ with entries in the selected order and let X be the X-matrix with the columns in this same order. Let $X^TX = C^TC$ be the Cholesky decomposition of X^TX, so that C is non-singular lower triangular. Then

$$g(x) = C^{-1}f(x)$$

is a vector of orthonormal polynomials with respect to D in the chosen order and we can write $g_\alpha(x), \alpha\in L$.

Now we attempt to generalise Gaussian quadrature in an informal fashion. Select some high degree orthonormal polynomial $g_\alpha,\ \alpha\in M\subset L$ and consider the solutions of the (simultaneous) set of equations

$$g_\alpha(x)=0,\ \alpha\in Q$$

and suppose this is a finite set, $D^* = \{z^{(1)},\dots z^{(q)}\}$. Repeat the construction of a saturated basis but with respect to D^*. Let the basis be $\{x^\alpha,\ \alpha\in L^*\}$ and suppose also that $L^*\subset L$.

The key step, as in one dimensional Gaussian quadrature, is to divide out a given polynomial $p(x)$ by the othonormal polynomial to obtain a remainder:

$$p(x) = \sum_{\alpha\in M} s_\alpha(x)g_\alpha(x) + R(x), \tag{12}$$

where

$$R(x) = \sum_{\alpha\in L^*}\theta_\alpha x^\alpha.$$

We needs one extra condition to give the quadrature:

(C): for each $\alpha \in M$ every monomial x^β in s_α has β lower down the order than α.

Assume (C) holds and integrate (12) with respect to design measure and use the orthogonality to obtain

$$\frac{1}{n}\sum_{x \in D} p(x) = \sum_{x \in D} R(x).$$

Just as for the one dimensional case, let the indicator function on D^* with respect to the L^* basis be $L_r(x)$, $(r = 1, \ldots, q)$ and write

$$R(x) = \sum_{r=1}^{q} p(z^{(r)}) L_r(x).$$

Then the quadrature is

$$\frac{1}{n}\sum_{x \in D} p(x) = \sum_{r=1}^{q} w_r p(z^{(r)})$$

where

$$w_r = \frac{1}{n}\sum_{x \in D} L_r(x).$$

The final step is to put $p(x) = x^\beta$ for those β required in our moment matching problem.

As an example we take $d = 2$ and the simple interaction model

$$\eta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2.$$

It can easily be shown by counter example that four independence points do not always exist in this case. We shall use the above method to obtain conditions on the design for four independence points to exist. Note that eight moments need to be matched on D and D^*:

$$\mu_{10}, \mu_{01}, \mu_{11}, \mu_{20}, \mu_{02}, \mu_{21}, \mu_{12}, \mu_{22}.$$

Suitable candidates for the set of orthonormal polynomials (M, above) are two quadratic polynomials with leading terms x_1^2 and x_2^2.

Let the order start

$$1, x_1, x_2, x_1^2, x_1 x_2, x_2^2, \ldots$$

This defines the order of the Cholesky decomposition. We quickly see a difficulty in trying to verify (6). Take the term $p(x) = x_1^2 x_2^2$. Suppose the orthonormal polynomial corresponding to x_2^2 is

$$q_{20}(x_1, x_2) = a_1 + a_2 x_1 + a_3 x_2 + a_4 x_1^2 + a_5 x_1 x_2 + a_6 x_2^2.$$

Then consider μ_{22} and write the appropriate monomial:

$$x_1^2 x_2^2 = x_1^2 q_{20}(x_1, x_2) + \ldots$$

The right hand side then has a term $a_4 x_1^4$. For (C) to hold we need $a_4 = 0$. This is a complex condition but after some algebra we find that sufficient conditions are:

$$\mu_{10} = \mu_{01} = \mu_{21} = \mu_{12} = \mu_{31} = \mu_{13} = \mu_{22} - \mu_{20}\mu_{02} = 0. \tag{13}$$

With these conditions we have up to a scalar multiple

$$\begin{aligned} q_{20} &= x_1^2 - \mu_{20}, \\ q_{02} &= x^2 - \mu_{02}. \end{aligned}$$

The solution to the simultaneous equations $q_{20} = q_{02} = 0$ gives the points

$$(z_1, z_2) = (\pm\sqrt{\mu_{20}}, \pm\sqrt{\mu_{02}}).$$

We see from the form of q_{20} and q_{02} that (C) is satisfied and the interpolation is possible. We can verify that with these conditions these are independence points that give (6).

Figure 2: Independence point band schematic: interaction model

It is not necessary for the design to have a factorial product structure. The "star composite" design

$$(\pm 1, \pm 1), (\pm c, 0), (0, \pm c)$$

satifies the conditions with $c = \sqrt{-2 + 2\sqrt{2}} = 0.9101...$ and then

$$(z_1, z_2) = \left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$$

Figure 2 shows the shape of the band. It consist of flat planar pieces around the origin and the x_1, x_2 axes and curvilinear pieces around the lines $x_1 = \pm x_2$.

4 Discussion

It has been shown that the independence point method extends to higher dimensions but that in general the existence of a complete set of k independence points may require conditions on the design. It is clear that a construction is intimately related to multivariate Gaussian quadrature with respect to the "design measure".

The constructions given here are not unique and in fact there is typically a family of solutions. For example in the linear regression case of Section 2 any Z with the property given in (6) provides a solution. It may be that different choices can be compared using extra criteria such as minimum average width.

When it is not possible to find k independence points we can sometimes find less than k independence points. Indeed for a given design there will be a maximum number of independence points. It is possible to give constructions of bands using less than a full set of independence points, but these bands are somewhat harder to interpret.

Finally, a note on construction. The plots in Figures 1 and 2 were obtain by simply computing all (appropriate) polynomials through the upper and lower ends of the interval at the independence points and evaluating the upper and lower envelopes using a *max* and *min* function, respectively (for this paper in MAPLE). It is possible, but somewhat lengthy, to ascertain mathematically exactly at which parts of the envelope which functions dominate. The shape in Figure 1 (seven flat pieces in each envelope) was computed before plotting but the rather beautiful 9-piece envelopes in Figure 2 was discovered by plotting first.

It seems a challenge to find independence points and study the envelopes for more complex cases such as general linear/interaction models.

References

Casella, G. and Strawderman, W.E. (1980), "Confidence bands for linear-regre -ssion with restricted predictor variables," *J. Amer. Stat. Assoc.*, 75, 862-868.

Gafarian, A. V. (1964), "Confidence bands in straight line regression," *Journal of the American Statistical Association*, 59, 182-213.

Graybill, F. A. and Bowden, D. C. (1967), "Linear Segment Confidence Bands for Simple Linear Regression Models," *Journal of the American Statistical Association*, 62, 403-408.

Hsu, J. C. (1996), "Multiple comparisons - theory and methods," Chapman & Hall.

Liu, W., Jamshidian, M., Zhang, Y. and Donnelly, J. (2004), " Simulation-based simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals," *J Computational and Graphical Statistics*, to appear.

Nainan, D. Q. (1986), "Conservative confidence bands in curvilinear regression," *Ann. Statist.*, 14, 896-906.

Naiman, D. Q. (1990), "Volumes of tubular neighborshoods of spherical polyhedra and statistical inference," *Ann. Statist.*, 18, 685-716.

Pan, W., Piegorsch, W.W. and West, R.W. (2003), " Exact one-sided simultaneous confidence bands via Uusipaikkaś method," *Ann. Instit. Statist. Math.*, 55 (2), 243-250.

Pistone, P., Riccomagno, R. and Wynn, H. P. (2000), "Algebraic Statistics," Chapman/CRC, Chichester.

Scheffe, H. (1953), "A method for judging all contrasts in the analysis of variance," *Biometrika*, 41, 153-169.

Scheffe, H. (1959), "The analysis of variance," Wiley.

Sun,J., Loader, C.R. and McCormick, W.P. (2000), "Confidence bands in generalized linear models," *Ann. Statist.*, 28, 429-460.

Tamhane, A. C. and Hochberg, Y. (1987), "Multiple comparison procedures," Wiley, New York.

Weyl, H. (1939), "On the volume of tubes," *Amer. J. Math. Soc.*, 61, 461-472.

Working, H. and Hotelling, H. (1929), "Applications of the theory of error to the interpretation of trends," *Journal of the American Statistical Association*, 24, 73-85.

Wynn, H. P. and Bloomfield, P. (1971), "Simultaneous confidence bands in regression (with discussion)," *J. Roy. Statist. Boc.*,B, 33, 202-217.

Wynn, H. P. (1984), "An exact confidence band for one-dimensional polynomial regression," *Biometrika*, 71, 375-9.