In a standard linear model, the test statistics

$$F = \frac{(RSS_0 - RSS_1)/(df_1 - df_0)}{RSS_1/df_1}$$

[also equivalent to comparing fitted values]

$$\frac{1}{q\hat{\sigma}^2}\hat{\underset{\sim}{\theta}}_q^T A^T \left[A(X^TX)^{-1}A^T\right]^{-1} A\hat{\underset{\sim}{\theta}}_q$$

[where A picks out the terms of interest.]

See Sobert, pp. 96-97.

The second form is convenient because it doesn't require refitting the model.

The motivation is that $A\hat{\underset{\sim}{\theta}}$ has variance

$$\operatorname{var}(A\hat{\theta}) = \operatorname{var}\left(A(X^TX)^{-1}X^T\underset{\sim}{y}\right) = \sigma^2 A(X^TX)^{-1}X^TX(X^TX)^{-1}A^T$$
$$= \sigma^2 A(X^TX)^{-1}A^T.$$

For p-splines, $A\hat{\underset{\sim}{\alpha}}$ has variance

$$\operatorname{var}(A\hat{\alpha}) = \operatorname{var}(A.BI.B^T y) = \sigma^2 A.BI.(B^TB)BI^TA^T$$

So, the test statistic is

$$\frac{1}{q\hat{\sigma}^2} \cdot \underset{\sim}{\alpha}^T A^T.\left[A\,BI.(B^TB)BI^TA^T\right]^{-1}.A\underset{\sim}{\alpha}$$

Since B^TB may not be of full rank the matrix being inverted in the middle of this expression may not be of full rank. An approximation may therefore be required.

Numerator is : $y^T B . B_1 . A^T . A_0^{-1} . A B_1 B^T \underset{\sim}{y}$

$$= y^T B . \left(B_1 A^T . A_0^{-1} A B_1 \right) . B^T y$$

$$= y^T B . A_1 . B^T y$$ ~~(the new notation)~~.

Denominator is : $RSS = y^T (I-P)^T (I-P) \underset{\sim}{y}$

$$= y^T \left(I - B B_1 B^T \right)^T \left(I - B B_1 B^T \right) \underset{\sim}{y}$$

$$= y^T \left(I - B \left(2 . B_1 - B_1 (B^T B) B_1 \right) B^T \right) \underset{\sim}{y}$$

$$= \underset{\sim}{y}^T \left(I - B A_2 B^T \right) \underset{\sim}{y}$$

$$P_r \{ F > F_{obs} \} = P_r \left\{ \frac{y^T B A_1 B^T y}{y^T (I - B A_2 B^T) \underset{\sim}{y}} > F_{obs} \right\}$$

$$= P_r \left\{ y^T \left[B A_1 B^T - F_{obs} (I - B A_2 B^T) \right] y^T > 0 \right\}$$

$$= P_r \left\{ \underset{\sim}{y}^T \left[B \left(A_1 + T_{obs} . A_2 \right) B^T - T_{obs} I \right] \underset{\sim}{y}^T > 0 \right\}$$

$$= P_r \left\{ \underset{\sim}{y}^T \left[B . A_3 B^T - T_{obs} . I \right] y^T > 0 \right\}$$

$A_3 = A_1 + T_{obs} . A_2$

$A = B . A_3 B^T - T_{obs} . I$

$tr\{A\} = tr\{B^T B . A_3\} - n T_{obs}$.

$$\cancel{\mathrm{tr}\{A^2\}} = \cancel{\mathrm{tr}\{\ \}}$$

$$A^2 = \left(B \cdot A_3 B^T - T_{OBS} \cdot I\right)\left(B A_3 B^T - T_{OBS} \cdot I\right)$$

$$= B A_3 \left(B^T B\right) A_3 B^T - 2\, T_{OBS}\, B A_3 B^T + T_{OBS}^2 \cdot I$$

$$= B \left[A_3 \left(B^T B\right) A_3 - 2\, T_{OBS}\, A_3 \right] B^T + T_{OBS}^2 \cdot I$$

$$= B \cdot A_4 \cdot B^T + T_{OBS}^2 \cdot I$$

$$\mathrm{tr}\{A^2\} = \mathrm{tr}\{B^T B \cdot A_4\} + n \cdot T_{OBS}^2 .$$

$$A^3 = \left(B \cdot A_4 \cdot B^T + T_{OBS}^2\, I\right)\left(B \cdot A_3 \cdot B^T - T_{OBS} \cdot I\right)$$

$$= B \cdot A_4 \left(B^T B\right) A_3 B^T + T_{OBS}^2\, B A_3 B^T$$

$$- T_{OBS} \cdot B \cdot A_4 B^T - T_{OBS}^3 \cdot I$$

$$= B \left[A_4 \left(B^T B\right) A_3 + T_{OBS}^2\, A_3 - T_{OBS}\, A_4 \right] B^T - T_{OBS}^3 \cdot I$$

$$= B \cdot A_5 \cdot B^T - T_{OBS}^3 \cdot I$$

$$\mathrm{tr}\{A^3\} = \mathrm{tr}\{B^T B \cdot A_5\} - n \cdot T_{OBS}^3 .$$

$H_0: \quad A\underset{\sim}{\alpha} = \underset{\sim}{0}.$

Numerator:

$$\underset{\sim}{\hat{\alpha}}^T A^T A_0^{-1} A \underset{\sim}{\hat{\alpha}} \qquad\qquad \underset{\sim}{\hat{\alpha}} = B_1 B^T \underset{\sim}{y}$$

$$\underset{\sim}{y}^T B B_1 A^T A_0^{-1} A B_1 B^T \underset{\sim}{y}$$

$$A\underset{\sim}{\hat{\alpha}} = A B_1 B^T y = \underbrace{A B_1 B^T B \underset{\sim}{\alpha}}_{(*)} + A B_1 B^T B \underset{\sim}{\varepsilon}$$

The terms involved in $A\underset{\sim}{\alpha}$ are all 0, so the penalty term inside B_1 should be small for those terms. For the part of $B_1 \cdot B^T B$ pulled out by A, we should therefore expect B_1^{-1} to be close to $B^T B$.

The component $(*)$ should therefore be close to 0. We might then reasonably expect $A\underset{\sim}{\hat{\alpha}}$ to behave like $A B_1 B^T B \underset{\sim}{\varepsilon}$.

$$\hat{\alpha} = (B^T B + \lambda)^{-1} B^T \underset{\sim}{y} = BI . B^T y$$

$$var(\hat{\alpha}) = BI . (B^T B + \lambda) BI . \sigma^2$$

$$\underset{\sim}{\hat{\alpha}}^T \left(BI (B^T B) (BI) \right)^{-1} \underset{\sim}{\alpha} = \underset{\sim}{y}^T B BI . BI^{-1} (B^T B)^{-1} BI^{-1} BI B^T \underset{\sim}{y}$$

Problem, because not necessarily invertible.

TEST STATISTIC

$$\hat{\underset{\sim}{\alpha}}_s^T \left[J_s \left[\underset{(\text{cov.alpha})}{B1 . \text{btb} . t(B1)} \right] J_s^T \right]^{-1} \hat{\underset{\sim}{\alpha}}_s \qquad J_s : m \times p.$$

$$= \hat{\underset{\sim}{\alpha}}^T J_s^T \left[J_s . \text{cov.alpha} . J_s^T \right]^{-1} J_s \hat{\underset{\sim}{\alpha}}$$

$$= \underset{\sim}{y}^T B B1^T . J_s^T \left[J_s . \text{cov.alpha} . J_s^T \right]^{-1} J_s B1 . B^T . \underset{\sim}{y}$$

$$= \underset{\sim}{y}^T B A_1 B^T \underset{\sim}{y}$$

Denominator is $(I - B A_2 B^T)$.

$$A = B A_1 B^T - t_{obs} . (I - B A_2 B^T)$$

Results derived earlier then follow.

AMEND

samsi

NSF • Duke • NCSU • UNC • NISS

$$\underset{\sim}{\alpha}_s^T \underset{\sim}{\alpha}_s = \underset{\sim}{y}^T B . B1 . I_s . B1 . B^T \underset{\sim}{y}$$

$$\hat{\underset{\sim}{\alpha}}_s^T B^T B \hat{\underset{\sim}{\alpha}}_s = y^T B . B1 . I_s (B^T B) I_s . B1 . B^T . \underset{\sim}{y} = \underset{\sim}{y}^T B . \underset{p\times p}{A_1} . B^T \underset{\sim}{y}$$

$$(I-P)^T(I-P) = \left(I - B.B1.B^T\right)^T\left(I - B.B1.B^T\right)$$

$$= \left(I - B B1 . B^T\right)\left(I - B.B1.B^T\right)$$

$$= \left(I - 2 B . B1 . B^T + B . B1 (B^T B) . B1 B^T\right) = I - B\left(2 B1 - B1 . (B^T B) B1\right.$$

$$= I - B . A_2 B^T .$$

$$A = B B1 . I_s . (B^T B) . I_s . B1 . B^T - t_{OBS} . \left(I - 2 B . B1 . B^T + B . B1 . (B^T B) B1 . B^T\right.$$

$$= B . B1 . I_s . (B^T B) . I_s . B1 . B^T - t_{OBS} . B . B1 (B^T B) . B1 . B^T - t_{OBS}\left(I - 2 B . B1 .\right.$$

$$= B . B1 . \left(I_s (B^T B) . I_s - t_{OBS} . (B^T B)\right) B1 . B^T - t_{OBS}\left(I - 2 B . B1 . B^T\right.$$

$$tr\{A\} = tr\left\{(B^T B) . \left[B1 . I_s (B^T B) I_s - t_{OBS} (B^T B)\right] . B1\right\} - t_{OBS}\left(n - 2 (B^T B) . B1\right.$$

$$A = B A_1 . B^T - t_{OBS} . \left(I - B . A_2 B^T\right)$$

$$= B(\underbrace{A_1 + t_{OBS} A_2}_{A_3}) B^T - t_{OBS} . I = B . A_3 . B^T - t_{OBS} . I$$

$$tr\{A\} = tr\{B^T B . A_3\} - n . t_{OBS}$$

$$A^2 = \left(B . A_3 B^T - t_{OBS} . I\right)\left(B A_3 B^T - t_{OBS} I\right)$$

$$= B . A_3 (B^T B) . A_3 B^T - 2 t_{OBS} B A_3 B^T + t_{OBS}^2 . I$$

$$= B . \left[A_3 (B^T B) A_3 - 2 t_{OBS} . A_3\right] B^T + t_{OBS}^2 . I$$

$$= B . A_4 B^T + t_{OBS}^2 . I$$

$$\mathrm{tr}\{A^2\} = \mathrm{tr}\{B^T B . A_4\} + n . t_{OBS}^2$$

$$A^3 = \left(B . A_4 B^T + t_{OBS}^2 I\right)\left(B A_3 B^T - t_{OBS} . I\right)$$

$$= B A_4 (B^T B) A_3 B^T + t_{OBS}^2 B A_3 B^T - t_{OBS} B A_4 B^T - t_{OBS}^3 . I$$

$$= B . \left[A_4 (B^T B) A_3 + t_{OBS}^2 A_3 - t_{OBS} . A_4\right] B^T - t_{OBS}^3 . I$$

$$= B A_5 . B^T - t_{OBS}^3 I$$

$$\mathrm{tr}\{A^3\} = \mathrm{tr}\{B^T B . A_5\} - n . t_{OBS}^3 .$$

Put this into an anova.sm function.

SAMSI
NSF • Duke • NCSU • UNC • NISS

$$\operatorname{tr}\{A\Sigma\} = \operatorname{tr}\{(BA_3B^T - t_{OBS} I)\Sigma\} = \operatorname{tr}\{BA_3B^T\Sigma - t_{OBS}\Sigma\}$$

$$= \operatorname{tr}\{B^T\Sigma B . A_3\} - t_{OBS} \operatorname{tr}\{\Sigma\}$$

etc.

Can we efficiently calculate $B^T\Sigma B$?

$$\underset{\sim}{y} = B\underset{\sim}{\theta} + \underset{\sim}{\varepsilon}$$

Minimise $\quad (\underset{\sim}{y} - B\underset{\sim}{\theta})^T(\underset{\sim}{y} - B\underset{\sim}{\theta}) + \lambda \underset{\sim}{\theta}^T D^T D \underset{\sim}{\theta}$

$$= \underset{\sim}{y}^T\underset{\sim}{y} - 2\underset{\sim}{\theta}^T B^T \underset{\sim}{y} + \underset{\sim}{\theta}^T B^T B \underset{\sim}{\theta} + \lambda \underset{\sim}{\theta}^T D^T D \underset{\sim}{\theta}$$

$$= \underset{\sim}{y}^T\underset{\sim}{y} + \underset{\sim}{\theta}^T (B^T B + \lambda D^T D)\underset{\sim}{\theta} - 2\underset{\sim}{\theta}^T B^T \underset{\sim}{y}$$

$$= \underset{\sim}{y}^T\underset{\sim}{y} + \left[\theta - (B^T B + \lambda D^T D)^{-1} B^T \underset{\sim}{y}\right]^T (B^T B + \lambda D^T D) \left[\; - \; \cdot \; \cdot \; \right]$$

$$= \cancel{\underset{\sim}{y}^T\underset{\sim}{y}} + \underset{\sim}{y}^T\underset{\sim}{y} - \underset{\sim}{y}^T B (B^T B + \lambda D^T D)^{-1} B^T \underset{\sim}{y}$$

$$\text{when} \quad \hat{\underset{\sim}{\theta}} = (B^T B + \lambda D^T D)^{-1} B^T \underset{\sim}{y}$$

$$\cancel{= (\underset{\sim}{y} - B(B^TB+\lambda D^TD)^{-1}B^T\underset{\sim}{y})^T(\underset{\sim}{y} - B(B^TB+\lambda D^TD)^{-1}B^T\underset{\sim}{y}) + [\theta]}$$

Under the constraint that $A\underset{\sim}{\beta} = \underset{\sim}{c}$

$$\hat{\underset{\sim}{\beta}}_H = \hat{\underset{\sim}{\beta}} - \tfrac{1}{2}(B^T B + \lambda D^T D)^{-1} A^T \hat{\lambda}_H$$

$$\underset{\sim}{c} = A\hat{\beta} - \tfrac{1}{2} A (B^T B + \lambda D^T D)^{-1} A^T \hat{\lambda}_H$$

$$\therefore \; -\tfrac{1}{2}\hat{\lambda}_H = \left[A (B^T B + \lambda D^T D)^{-1} A^T\right]^{-1} (\underset{\sim}{c} - A\hat{\beta})$$

So

$$\hat{\underset{\sim}{\beta}}_H = \hat{\beta} + (B^T B + \lambda D^T D)^{-1} A^T \left[A (B^T B + \lambda D^T D)^{-1} A^T\right]^{-1} (\underset{\sim}{c} - A\hat{\beta}).$$

Multiple penalties, if required.

(A ~~is~~ is non-zero only in a block on the diagonal.

(A is second-difference matrix for linearity.))

INFERENCE FOR P-SPLINE ADDITIVE MODELS

TEST OF NO EFFECT

Possible test statistics are

$$\frac{(RSS_0 - RSS_1)/(df_1 - df_0)}{RSS_1/(n - df_1)}$$

$$\frac{\hat{\beta}^T \cdot \hat{\Sigma}_{\hat{\beta}}^{-1} \hat{\beta} / \cancel{\text{normalised}}\,(df_1 - df_0)}{RSS_1/(n - df_1)}$$

$$\frac{\sum_{i=1}^{n} (\hat{m}_{i1} - \hat{m}_{i0})^2 / (df_1 - df_0)}{RSS_1/(n - df_1)}$$

These statistics are all equivalent for linear models (Searle, Theorem 4.1), but may not be for additive models.

In other words, $RSS_0 - RSS_1$ can be represented as $\hat{\beta}^T \cdot \hat{R}_{\hat{\beta}} \hat{\beta}$ or as $\sum_{i=1}^{n} (\hat{m}_{i1} - \hat{m}_{i0})^2$.

The $\hat{\beta}_H^T R_{\hat{\beta}} \hat{\beta}_H$ form is attractive because of its lower dimensionality.

$$E = \begin{bmatrix} \underset{\sim}{e}_1 & \cdots & \underset{\sim}{e}_m \end{bmatrix} \qquad E^T = \begin{pmatrix} \underset{\sim}{e}_1^T \\ \vdots \\ \underset{\sim}{e}_m^T \end{pmatrix}$$

$$(E E^T]_{ij} = \sum_k e_{ik}$$

$$\Sigma = Q \Lambda Q^T$$

$$\Sigma^{-1} = Q \Lambda^{-1} Q^T$$

$$\Sigma \Sigma^{-1} = Q \Lambda Q^T Q \Lambda^{-1} Q^T$$

$$= Q Q^T.$$

$$= I \quad \text{since } Q^{-1} = Q^T.$$

$$\hat{\alpha} = B1.\, B^T \underset{\sim}{y}$$

$$\text{cov}(\hat{\alpha}) = B1.\, B^T.\, \sigma^2\, B\, B1$$

$$B1 = (B^TB + \lambda\, \Delta^T\Delta)^{-1} \quad : \text{ symmetric.}$$

However, $\text{cov}(\hat{\alpha})$ is not of full rank.
(see Wood p.195, top).

So, use other test statistics. We don't need to refit the model because of the adjustment formulae.

But these all involve the inverse of $\text{cov}(\hat{\alpha})$.

Pseudo-inverse:
$$\hat{\Sigma} = \sum_{i=1}^{m} \lambda_i\, \underset{\sim}{e}_i\, \underset{\sim}{e}_i^T$$

$$\hat{\Sigma}^{-1} = \sum_{i=1}^{m} \frac{1}{\lambda_i}\, \underset{\sim}{e}_i\, \underset{\sim}{e}_i$$

$$\hat{\Sigma}\hat{\Sigma}^{-1} = \sum_i \sum_j \lambda_i \frac{1}{\lambda_j}\, \underset{\sim}{e}_i\, \underset{\sim}{e}_i^T\, \underset{\sim}{e}_j\, \underset{\sim}{e}_j^T$$

$$= \sum_i \underset{\sim}{e}_i\, \underset{\sim}{e}_i^T = E\, E^T$$

$$= \sum_i \sum_j \left[e_{ji}\, e_{j'i} \right]_{jj'}$$

When A is ~~diag((0, 1, 0, 1, ..., 1, 0, ..., 0))~~

$m \times n$ $\begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & & & \\ 0 & & & & 0 & 1 & 0 & \cdots & & \\ & & & & & & \ddots & & & \\ & & & & & & & 1 & 0 & \cdots \end{pmatrix}$

BIAS

$$y_i = m(x_i) + \varepsilon_i$$

$$\hat{\underset{\sim}{m}} = B(B^TB + \lambda D^TD)^{-1} B^T \underset{\sim}{y}$$

$$E(\hat{\underset{\sim}{m}}) = B(B^TB + \lambda D^TD)^{-1} B^T \underset{\sim}{m} = B\left(B^TB\left[I + \lambda(B^TB)^{-1}(D^TD)\right]\right)^{-1} B^T \underset{\sim}{m}$$

$$= B\left[I + \lambda(B^TB)^{-1}(D^TD)\right]^{-1}(B^TB)^{-1}B^T \underset{\sim}{m}$$

~~If $\underset{\sim}{m}$ has a representation as $B\underset{\sim}{\alpha}$ then this reduces to a simple form of ad~~

There are two adjustments here. One is the approximation of $\underset{\sim}{m}$ by $B\underset{\sim}{\alpha}$.

The other is the adjustment in the $[\ ]^{-1}$ term. This is a form which has an alternative analytic expression. Check this out.

Opsomer et al. (2009) Bka give asymptotic expressions. Krivobokova et al. for mixed model case.

TESTING Testing for $m(x) \equiv 0$.

This is a standard regression model, so compare RSS's.

$$RSS_1 = y^T(I - B(B^TB + \lambda D^TD)^{-1}B^T)^T(I - B(B^TB + \lambda D^TD)^{-1}B^T)y$$

$$= y^T\left[I - 2B(B^TB + \lambda D^TD)^{-1}B^T + B(B\ \ )^{-1}B^TB(\ \ )^{-1}B^T\right]\underset{\sim}{y}$$

$$RSS_0 = y^T[\ \ \dots\ \ B_0\ \ ]\underset{\sim}{y}$$

$$(RSS_0 - RSS_1) = y^T\left[(I - P_0)(I - P_0) - (I - P_1)(I - P_1)\right]\underset{\sim}{y}$$

$$= \underset{\sim}{y}^T\left[P_0^2 - 2P_0 - P_1^2 - 2P_1\right]\underset{\sim}{y} = \underset{\sim}{y}^T A \underset{\sim}{y}$$

The quadratic form argument is based on $\mathrm{tr}\{A^j\}$.

$\mathrm{tr}\{A\}$ is trace of a $p \times p$ matrix.

$\mathrm{tr}\{A^2\}$ involves P_0P_1 & P_1P_0.

$$P_0P_1 = B_0(B_0^TB_0 + \lambda D_0^TD_0)^{-1}B_0^T . B_1(B_1^TB_1 + \lambda D_1^TD_1)^{-1}B_1^T$$

This has lower dimension, so OK. $\boxed{\text{Look at Wood book for practice there.}}$

$$(I + A)^{-1} = I - A(I + A)^{-1}$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

$$\left[I + \lambda (B^TB)^{-1}(D^TD)\right]^{-1} = I - (B^TB + \lambda D^TD)^{-1}(D^TD)$$

P-GAMs

For each component, compute the model which is restricted to be linear, using the adjustment formula in Seber (p. 55) (Check this remains valid in the penalty world: "fixed" argument too.)

$$\left\| B\left(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}\right)\right\| = \left(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}\right)^T B^T B \left(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}\right)$$

$$= \left(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}_H + \hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right)^T B^T B \left(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}_H + \hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right)$$

$$= \left(\underset{\sim}{\theta} - \hat{\underset{\sim}{\theta}}_H\right)^T B^T B \left(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}_H\right) + \left(\hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right)^T B^T B \left(\hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right)$$

since

$$2\left(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}_H\right) B^T B \left(\hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right) = \hat{\lambda}_H^T A \left(B^T B + A D^T D\right)^{-1} B^T B \left(\hat{\underset{\sim}{\theta}}_H - \underset{\sim}{\theta}\right)$$

$$\underset{\sim}{\varepsilon}' \underset{\sim}{\varepsilon} = \left(\underset{\sim}{y} - X \underset{\sim}{\theta}\right)^T \left(\underset{\sim}{y} - X \underset{\sim}{\theta}\right)$$

$$= \left(\underset{\sim}{y} - X \hat{\underset{\sim}{\theta}}\right)^T \left(\underset{\sim}{y} - X \hat{\underset{\sim}{\theta}}\right)$$

P-GAMS

Innovation:

Est^ of σ^2 ~~[illegible]~~ may be improved by using a very small bandwidth. This can go in the denominator of F-statistics.

Is there a natural largest df (smallest λ) which can be used? ~~[illegible]~~ An upper ~~[illegible]~~ bound for df is min (n, nbases).

More exactly, the upper limit is the number of basis functions which have at least one non-zero evaluations ~~[illegible]~~ at the data points. If any basis functions has evaluations which are all zero then this upper limit cannot be achieved.

Anova by quadratic forms.

Explanation of properties by matrix inversion, when all basis functions have at least one non-zero evaluation.

Graphics for no-effect + linear terms, for 1-d + 2-d covariates.

Animation for 3-d covariates.

Addition of correlated errors? Simultaneous est^?

An alternative representation, which is equivalent because this is a regression model, is something like $\|\hat{\alpha}_s\|^2$, where $\hat{\alpha}_s$ is the subset of $\hat{\alpha}$ corresponding to the component of interest.

$$\hat{\alpha}_s = \left[(B^TB + \lambda D^TD)^{-1} B^T y\right]_s = \left[(BI)\cdot(B^Ty)\right]_s = (BI)_s B^T y$$

$$\|\hat{\alpha}_s\|^2 = y^T B (BI_s^T BI_s) B^T y$$, where BI_s has relevant rows set to 0.

There is probably a neater decomposition. (Done this before?) Look up Searle for details.

How does the bias argument work when a component is linear or 0?

---

sm code: infrastructure of regression should be usable. Local linear specific code is limited. | Better to have separate code + separate out plotting code.

Graphics for additive model components, including surfaces, could be added to the paper.

Possibly better in the long-run to do everything in terms of a general sm $f^\wedge$ for additive models?

Always put in an "x" column for each component so that the additional terms can be assessed for non-linearity? There may be an easier way to do this by assessing the $\hat{\alpha}_s$ terms appropriately: $\hat{\alpha}_s^T D^T D \hat{\alpha}_s = \|D\hat{\alpha}_s\|^2$. !!

Is there a case for using a large df denominator? ($\lambda = 0$). $\hat{\sigma}^2$ df is $\mathrm{tr}\{2P_1^2 - 2P_1\}$ which is different from Paul's $\mathrm{tr}\{P_1\}$.

Prototype

- things done better/faster (3 covariates slow)

~~[illegible]~~

Why

3 covariates

- understand
- do things in special ways (anova).

Developments

- GLAM
- link with sm infrastructure; options, graphics, panel
- correlated data.

Features

- df
- formula
- methods for plot, summary, anova, predict
- anova wrong but good
- plot of 3-d covariates.

Problem at x1, x2 section of additive models in sm-pam-test.r.

<u>sm.pgam</u>

Use that name?

Choose λ through df using univariate terms
(+ efficient use of $B^T B$). That's what
happens in the usual gam.

Fitted pgam has a projection matrix decomposition
just by taking the appropriate columns &
elements of $B\underset{\sim}{\alpha}$. Calculate df's from this.

Estimate error df & $\hat{\sigma}$.

Put all this in a summary f^n.

<u>Later</u>

anova f^n which looks at highest order terms
& performs tests.

Integrate with sm.

Allow linear terms.

DEGREES OF FREEDOM

All

$$B\hat{\alpha} = B\left(B^TB+P\right)^{-1}B^T\underset{\sim}{y}$$

$$df = tr\{B^TB\ B_1\}$$

Component

$$\hat{\underset{\sim}{m}}_i = B\,\underset{\sim}{\hat{\alpha}}_i = [0:B_i:0]\,\underset{\sim}{\hat{\alpha}}\ ,$$ ~~where B_i has 0's in cols. not assoc^d. with comp. i.~~

$$= [0:B_i:0]\left(B^TB+P\right)^{-1}B^Ty$$

$$P_i = [0:B_i:0]\,\cancel{(B^TB+P)}\,B_1\,B^T$$

$$tr\{P_i\} = tr\left\{B^T[0:B_i:0]\,.\,B_1\right\}$$

$$= tr\left\{\begin{bmatrix}B_a^T\\ B_i^T\\ B_b^T\end{bmatrix}[0:B_i:0]\,.\,B_1\right\}$$

$$= tr\left\{\begin{bmatrix}0 & B_a^TB_i & 0\\ 0 & B_i^TB_i & 0\\ 0 & B_b^TB_i & 0\end{bmatrix}B_1\right\} = tr\left\{B_i^TB_iB_1\right.$$

$$= tr\left\{B_i^TB_i\,B_{1i}\right\}$$

df for error?

R^2?

$$\boxed{\begin{aligned}&\underset{\sim}{\hat{\alpha}} = B1.B^Ty\\ &var(\underset{\sim}{\hat{\alpha}}) = B1.B^TB.B1^T.\sigma^2\\ &\cancel{RSS = y^T(I-BB1.B^T)^T(I-B1.B^T)y}\\ &df.error = \cancel{n - tr\{B.B1^T + B1.B^T - BB1.B1B^T\}}\end{aligned}}$$