Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

means + ineqalities + splines

Outline

Outline

Hypothesis tests of r=Rank(R) restrictions

- Equality restriction versus unconstrained

$$H_0 : R\theta = 0 \text{ versus } H_A : R\theta \neq 0$$

- Equality restrictions versus inequality restriction

$$H_0 : R\theta = 0 \quad \text{versus} \quad H'_A : R\theta \geq 0$$

- Inequality restricted vs unconstrained

$$H'_A : R\theta \geq 0 \quad \text{versus} \quad H'_B : R\theta \not\geq 0.$$

- Inequality restricted versus additional inequalities

$$H'_A : R_A\theta \geq 0 \quad \text{versus} \quad H'_B : R_B\theta \geq 0,\, R_A \subset R_B$$

Hypothesis tests of r=Rank(R) restrictions

- Fixed difference in dimensionality of models
  - Equality restriction versus unconstrained

$$Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c) = Pr(\chi^2_r > c)$$

- Difference in number of free parms is stochastic
  - Equality restrictions versus inequality restriction
  - Inequality restricted vs unconstrained
  - Inequality restricted versus additional inequalities

$$Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c) = Pr(\bar{\chi}^2 > c)$$
$$= \sum_{k=1}^{K} w_k Pr(\chi^2_k > c)$$

where w_k is the probability of having a difference of k degrees of freedom between models

Inference on a convex cone

Let

$$\hat{\Delta} = (\hat{\Delta}_1, \hat{\Delta}_2) \sim N(\Delta, I_2)$$

where, possibly,

$$\hat{\Delta}_j = \hat{\mu}_j - \hat{\mu}_{j-1} \qquad j \in \{1, 2\}$$

and I_k is a $k \times k$ identity matrix.

Hypotheses/comparisons

- $H_0 : \Delta = 0 \quad \text{vs} \quad H_{\emptyset} : \Delta \neq 0$
- $H_0 : \Delta = 0 \quad \text{vs} \quad H_{\nearrow} : \Delta \geq 0$

Inference on a convex cone

Let

$$\hat{\Delta} = (\hat{\Delta}_1, \hat{\Delta}_2) \sim N(\Delta, I_2)$$

where, possibly,

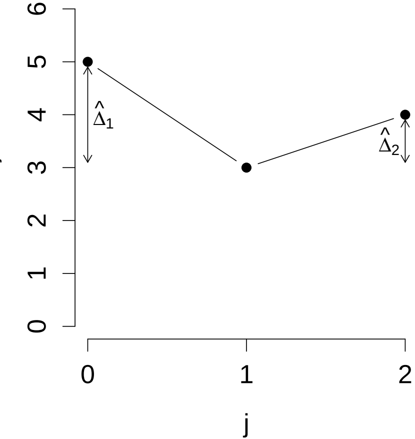
$$\hat{\Delta}_j = \hat{\mu}_j - \hat{\mu}_{j-1} \qquad j \in \{1, 2\}$$

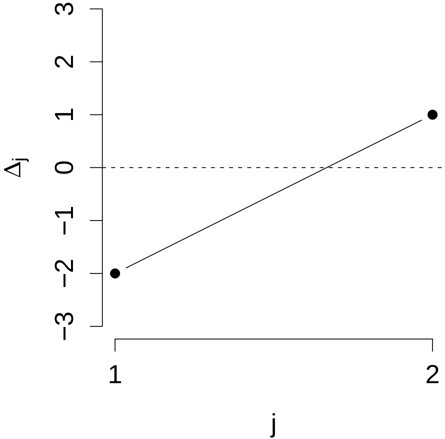
and I_k is a $k \times k$ identity matrix.

Hypotheses/comparisons

- $H_0 : \Delta = 0$ vs $H_{\emptyset} : \Delta \neq 0$
- $H_0 : \Delta = 0$ vs $H_{\nearrow} : \Delta \geq 0$

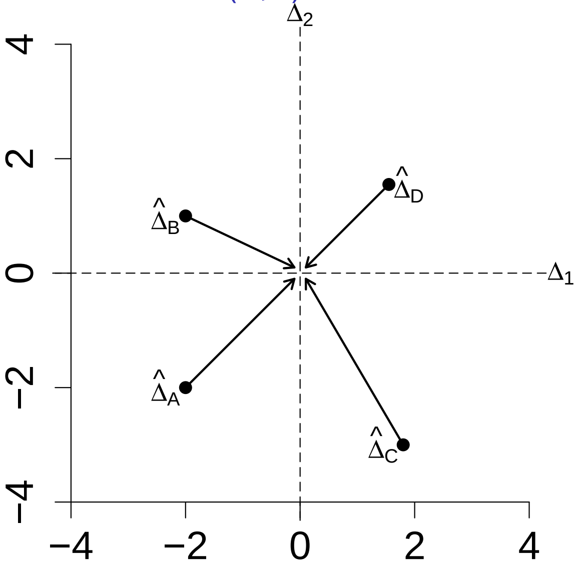
For example





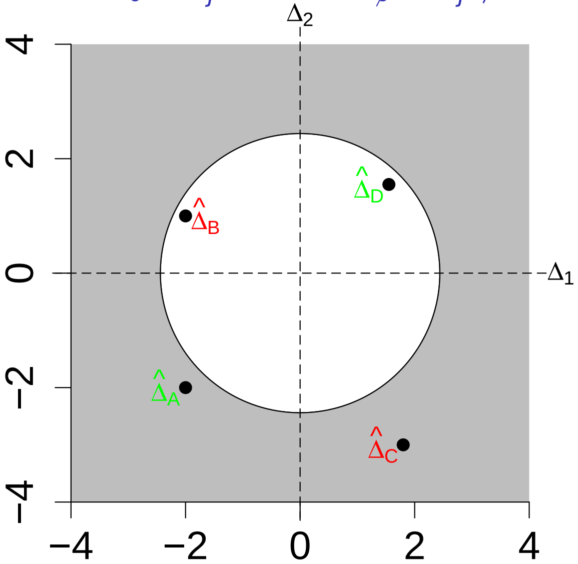
I will call this case "B", with notation $\hat{\mu}_B = (\hat{\mu}_{0B}, \hat{\mu}_{1B}, \hat{\mu}_{2B})$ and $\hat{\Delta}_B = (\hat{\Delta}_{1B}, \hat{\Delta}_{2B})$.

Distance between (0,0) and $\hat{\Delta}$



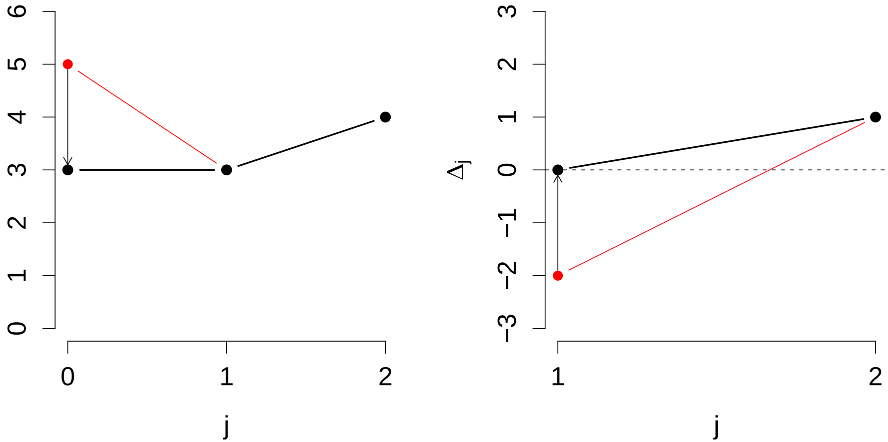
$\Delta_1^2 + \Delta_2^2 \sim \chi_2^2$

Critical set $H_0 : \Delta_j = 0$ vs $H_{\not 0} : \Delta_j \neq 0$



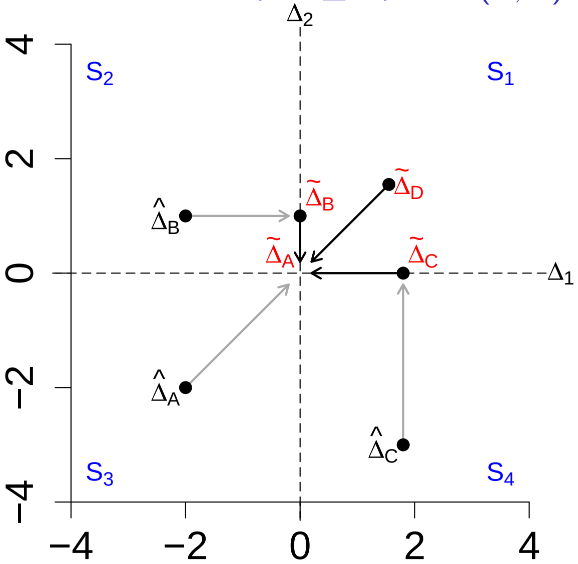
Critical value $\chi^2_{2,\alpha=.05} = 6$, so $\sqrt{\Delta_1^2 + \Delta_2^2} = \sqrt{6} =$ radius

Revisiting example B, imposing $H_{\nearrow}$ (monotonicity)



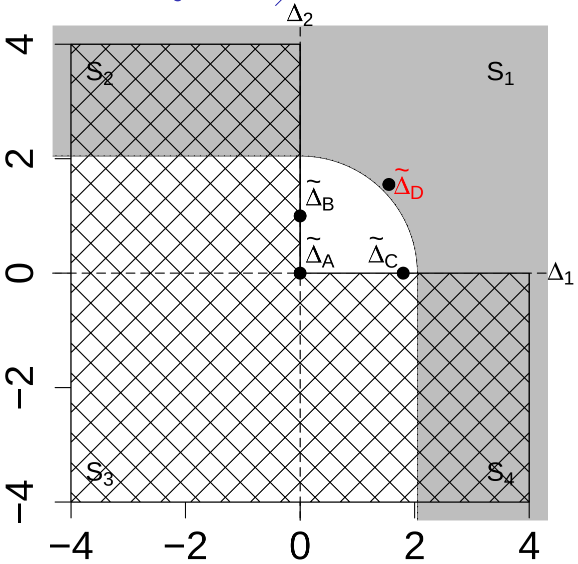
Red: unconstrained values changed to achieve monotonicity
Black dots are consistent with monotonic

Distance between $\hat{\Delta}$, $\tilde{\Delta} \geq 0$, and $(0, 0)$



$\tilde{\Delta}_A \sim 0 \quad \tilde{\Delta}_B, \tilde{\Delta}_C \sim \chi^2_1 \quad \tilde{\Delta}_D \sim \chi^2_2$

Critical set of H_0 vs $H_{\nearrow}$



$$P(\bar{\chi}^2_{+,\alpha} > \bar{c}) = 1/4P(\chi^2_2 > \bar{c}) + 1/2P(\chi^2_1 > \bar{c}) + 1/4 = \alpha$$

Constrained estimation: dimensions and distance

Constrained estimates $\tilde{\Delta}$, subject to $\Delta > 0$

$$\tilde{\Delta} = \begin{cases} (\hat{\Delta}_1, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (\hat{\Delta}_1, 0) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 < 0 \\ (0, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (0, 0) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 < 0 \end{cases}$$

How many free parameters? 2 (S1), 1 (S2, S4), or 0 (S3).

What is probability of being "far" from H_0? By quadrant:

$$P(\tilde{\Delta}_1^2 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_1) = P(\chi_2^2 < c')$$
$$P(\tilde{\Delta}_1^2 + 0 < c' \mid \hat{\Delta} \in S_2) = P(\chi_1^2 < c')$$
$$P(0 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_4) = P(\chi_1^2 < c')$$
$$P(0 + 0 < c' \mid \hat{\Delta} \in S_3) = 1$$

Constrained estimation: dimensions and distance

Constrained estimates $\tilde{\Delta}$, subject to $\Delta > 0$

$$\tilde{\Delta} = \begin{cases} (\hat{\Delta}_1, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (\hat{\Delta}_1, 0) & \text{if } \hat{\Delta}_1 \geq 0 \text{ and } \hat{\Delta}_2 < 0 \\ (0, \hat{\Delta}_2) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 \geq 0 \\ (0, 0) & \text{if } \hat{\Delta}_1 < 0 \text{ and } \hat{\Delta}_2 < 0 \end{cases}$$

How many free parameters? 2 (S1), 1 (S2, S4), or 0 (S3).
What is probability of being "far" from H_0? By quadrant:

$$\begin{aligned} P(\tilde{\Delta}_1^2 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_1) &= P(\chi_2^2 < c') \\ P(\tilde{\Delta}_1^2 + 0 < c' \mid \hat{\Delta} \in S_2) &= P(\chi_1^2 < c') \\ P(0 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_4) &= P(\chi_1^2 < c') \\ P(0 + 0 < c' \mid \hat{\Delta} \in S_3) &= 1 \end{aligned}$$

Constrained estimation: dimensions and distance

In this simple example, each of quadrants is equally likely under H_0

$$
\begin{aligned}
P(\tilde{\Delta}_1^2 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_1) &= P(\chi_2^2 < c') \\
P(\tilde{\Delta}_1^2 + 0 < c' \mid \hat{\Delta} \in S_2) &= P(\chi_1^2 < c') \\
P(0 + \tilde{\Delta}_2^2 < c' \mid \hat{\Delta} \in S_4) &= P(\chi_1^2 < c') \\
P(0 + 0 < c' \mid \hat{\Delta} \in S_3) &= 1
\end{aligned}
$$

So for given α, solve for c'

$$
P(\bar{\chi}^2 < c') = 1/4P(\chi_2^2 < c') + 1/2P(\chi_1^2 < c') + 1/4 = 1 - \alpha
$$

Distribution of hypothesis tests

- Equality restriction versus unconstrained
  - Fixed difference in dimensionality of models

$$Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c) = Pr(\chi^2_r > c)$$

- But number of free $-$ restricted parms is stochastic if
  - Equality restrictions versus inequality restriction
  - Inequality restricted vs unconstrained
  - Inequality restricted versus additional inequalities

$$Pr(-2[L(\tilde{\theta}) - L(\hat{\theta})] > c') = Pr(\bar{\chi}^2 > c') = \sum_{k=1}^{K} w_k Pr(\chi^2_k > c')$$

where w_k is the probability of having a difference of k degrees of freedom between models

Distribution of hypothesis tests

- For equality-based hypo., choose test size α, solve for c,

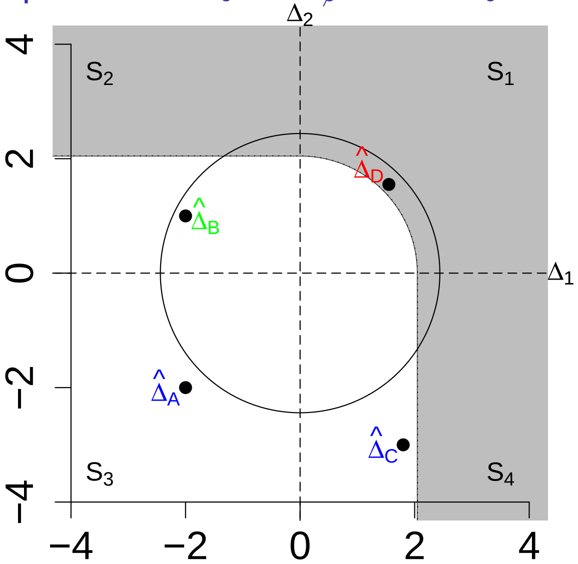
$$Pr(\chi^2_r > c) = \alpha$$

- For convex hypo., choose size α and solve for c',

$$\sum_{k=1}^{K} w_k Pr(\chi^2_k > c') = \alpha$$

- For any α, if $w_k < 1$ then $c' < c$
  Can use c as an upper bound (cf. Wand 2010)

Overlap of sets: H_0 vs $H_{\emptyset}$ and H_0 vs $H_{\nearrow}$



$$\sqrt{\chi^2_{2,\alpha}} = radius_{\emptyset} = 2.45 \quad \text{vs} \quad \sqrt{\bar{\chi}^2_{+,\alpha}} = radius_{\nearrow} = 2.05$$

Outline

functions of x

Let's think about aproximating $E(y|x) = f(x)$ by

$$f(x) = \sum_{m=0}^{M} \beta_m h_m(x)$$

Quadratic regression

$$h_0(x) = 1; \quad h_1(x) = x; \quad h_2(x) = x^2$$
$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

Broken stick

$$h_0(x) = 1; \quad h_1(x) = x; \quad h_2(x) = (x - .5)_+$$
$$f(x) = \beta_0 + \beta_1 x + \beta_2 (x - .5)_+$$

Sequence of means

$$h_m(x) = I(L_m \leq x < U_m)$$
$$f(x) = \beta_0 h_0 + \beta_1 h_1(x) + ... + \beta_m h_m(x)$$

functions of x

A whip like model,

$$h_0(x) = 1$$
$$h_1(x) = x$$
$$h_2(x) = (x - .5)_+$$
$$h_3(x) = (x - .55)_+$$
$$...$$
$$h_m(x) = (x - .95)_+$$

- this is a particular case of *linear spline basis function*
- piecewise linear
- "knot" locations:

$$\lambda = (.5, .55, ..., .95)$$

- Q: what does adding knots do to property of curve?
- Q: how do we pick knots?

functions of x

Q: what is the derivative of a function of linear basis at a knot?
Q: How might we get smoothness?

$$
\begin{aligned}
h_0(x) &= 1 \\
h_1(x) &= x \\
h_1(x) &= x^2 \\
h_2(x) &= (x - \lambda_1)_+^2 \\
&\ldots \\
h_m(x) &= (x - \lambda_K)_+^2
\end{aligned}
$$

- quadratic spline
- f has continuous first derivative at all points
- more generally, a tructated power basis of dgree p

$$(x - \lambda_k)_+^p$$

- has continuous $p - 1$ derivatives

Outline

Basis splines

Logic:

- map $x_i \rightarrow h_m(x_i)$ (x_i into M basis functions)
- estimate $f(x)$ (curve, a weighted sum of $h_m(x)$):

$$\hat{f}(x_i) = \sum_{m=1}^{M} \hat{\beta}_m h_m(x_i) = \beta h(x_i)$$

where $\hat{\beta} = \left[\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m\right]$ are simply regression coefficients.

Features of B-splines

- shapes can be described by linear functions of β
- $h_m(x)$ has local support, β_m has local effect

Basis splines, basic logic

- Given a choice of knot locations $\{\lambda_1, ..., \lambda_M\}$ and polynomial order k
- Decompose x_i into $M + 2$ basis functions, with mth

$$h_{m,k+1}(x) = (\lambda_{m+k+1} - \lambda_i) \sum_{j=0}^{k+1} \frac{(\lambda_{m+j} - x)_+^k}{\prod_{l=0, l \neq j}^{k+1} (\lambda_{m+j} - \lambda_{m+l})}$$

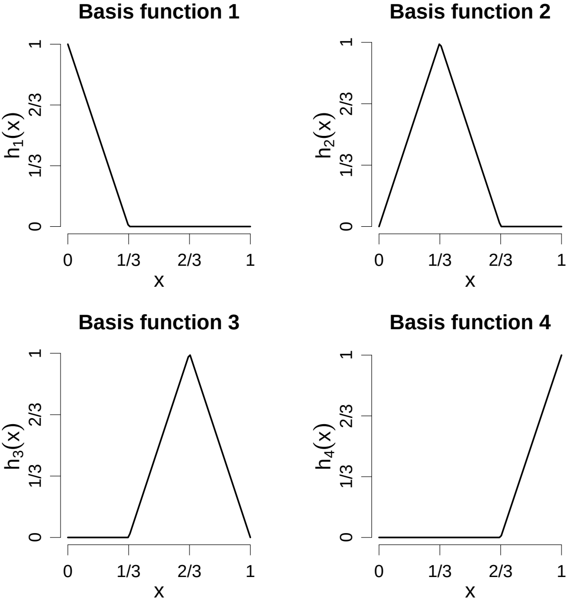
- for a vector of spline coefficients,

$$\hat{\beta} = \left[\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m\right]$$

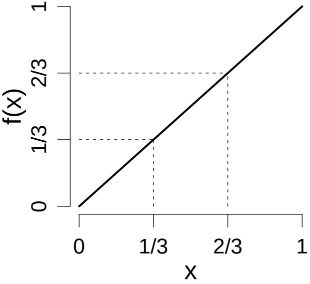
- $f(x)$ is a weighted sum of $h_m(x)$

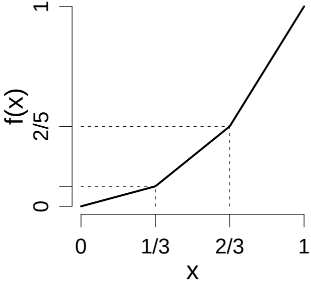
$$\hat{f}(x_i) = \sum_{m=1}^{M} \hat{\beta}_m h_m(x_i)$$

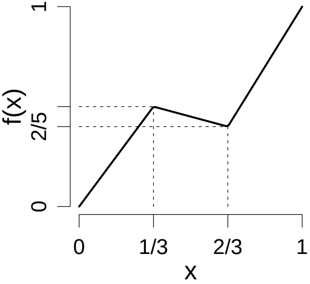
A look at h(x) of order 1, knots at 1/3 and 2/3



Mapping x (horizontal axis) to $f(x)$ (vertical axis)







(a) $\theta = [\,0,\ .33,\ .66,\ 1]$

$\Delta\theta = [\,.33,\ .33,\ .33]$

(b) $\theta = [\,0,\ .1,\ .4,\ 1]$

$\Delta\theta = [\,1,\ .3,\ .6]$

(c) $\theta = [\,0,\ .5,\ .4,\ 1]$

$\Delta\theta = [\,.5,\ -.1,\ .6]$

Linear constraints implying shapes

	Restriction	$f(x)$	*Interval*
1)	$\beta_m - \beta_{m-1} > 0$	increasing	(k_{m-1}, k_m)
2)	$\beta_m - \beta_{m-1} = 0$	flat	(k_{m-1}, k_m)
3)	$\beta_m - \beta_{m-1} < 0$	decreasing	(k_{m-1}, k_m)
4)	$\frac{\beta_{m+1} - \beta_m}{k_{m+1} - k_m} = \frac{\beta_m - \beta_{m-1}}{k_m - k_{m-1}}$	linear	(k_{m-1}, k_{m+1})
5)	$\frac{\beta_{m+1} - \beta_m}{k_{m+1} - k_m} > \frac{\beta_m - \beta_{m-1}}{k_m - k_{m-1}}$	convex	(k_{m-1}, k_{m+1})
6)	$\frac{\beta_{m+1} - \beta_m}{k_{m+1} - k_m} < \frac{\beta_m - \beta_{m-1}}{k_m - k_{m-1}}$	concave	(k_{m-1}, k_{m+1})

and can combine, e.g., monotonic and convex; unimodal

Linear constraints implying B-spline shapes

Linear restrictions on parameters, can be written as,

$$R\beta - c \geq 0$$

Example 1: monotonicity ($\beta_m - \beta_{m-1} > 0$)

$$R_m = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix},\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix},\ c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Example 2: symmetric

$$R_m = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix},\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix},\ c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Estimation of spline coefficients

- Unconstrained shape, $f(x)$ is linear function of β

$$\min \sum_{i}^{N} (y_i - h(x_i)\hat{\beta})^2$$

- Constrained OLS: quadratic programing problem

$$\min \sum_{i}^{N} (y_i - h(x_i)\tilde{\beta})^2 \quad \text{subject to } R\tilde{\beta} - c \geq 0$$

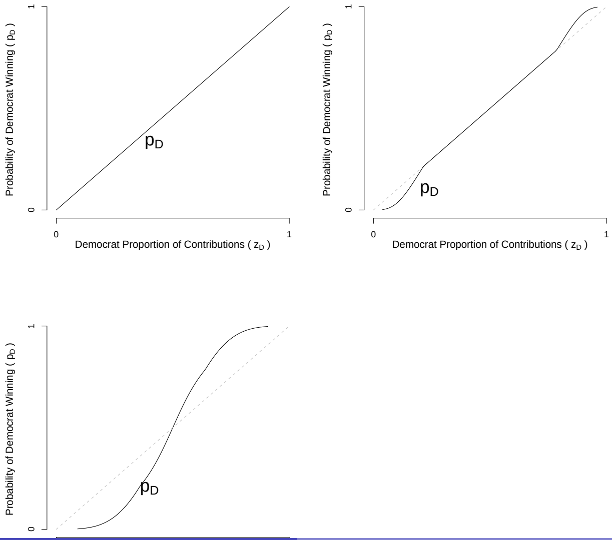
- Constrained, non-linear/ML: logarithmic barrier

$$\sum_{i}^{N} L(h(x_i)\tilde{\beta}) - \mu \sum \log(R\tilde{\beta} - c)$$

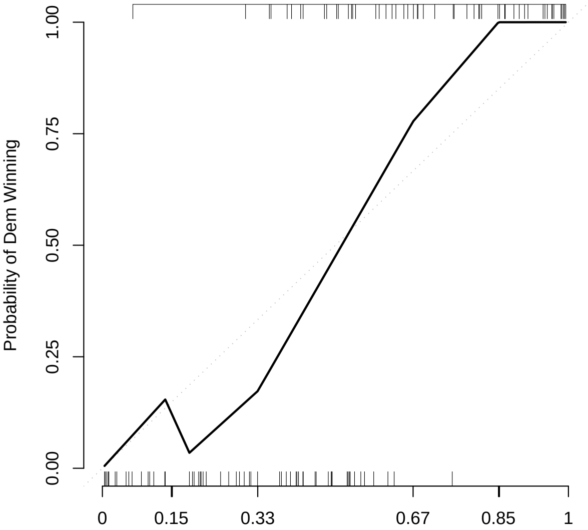
Outline

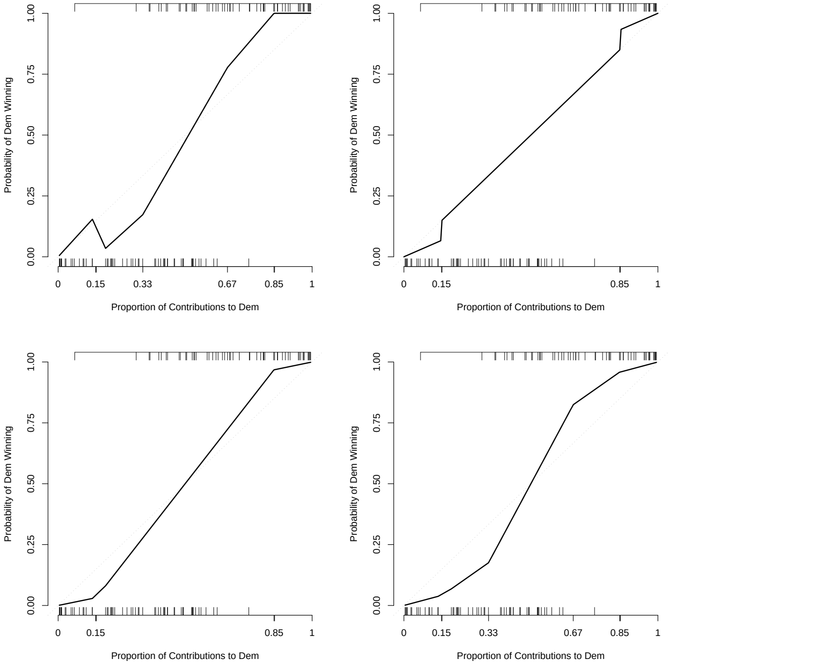
Motivations of PACs (Wand 2011)

Probability of Dem. victory by share of Dem. contributions



Open seats 1980–1986, unrestricted B-spline




Probability of Dem Winning
0.00
0.25
0.50
0.75
1.00
0
0.15
0.33
0.67
0.85
1
Proportion of Contributions to Dem

PAC motives: model comparisons

			$\Pr(\bar{\chi}^2 > c)$		
Model (j):	Parms m max	Log-lik. L_j	j vs Linear	j vs Dips	j vs Mono.
Linear Equality	0	-47.03			
w/ symmetric dips	1	-46.59	0.18		
Symmetric, monotonic	2	-44.44	0.06	0.03	
w/ knots at $(\frac{1}{3}, \frac{2}{3})$	3	-43.81	0.07	0.04	0.48
Unrestricted	6	-42.92	0.22	0.89	0.34
w/ knots at $(\frac{1}{3}, \frac{2}{3})$	8	-42.46	0.99	0.89	0.34

Note: $\bar{\chi}^2 = -2(L_{\text{row}} - L_{\text{column}})$

Concluding comments

Outline

Beyond average differences
... and arbitrary flexibleness

- Have a theory, ideally more than one
- "Make your theories elaborate" (Fisher / Cochran 1965):
  - when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible
  - if a hypothesis predicts that y will increase steadily as the causal variable z increases, a study with at least three levels of z gives a more comprehensive check than one with two levels
  - i.e, check shape! not just average change
- And check against omnibus alternatives
  but be clear this is for idea generation and robustness!

GLM extensions

Simple extension, back-fitting

- Bachetti (1989) Additive Isotonic Models
- Geyer, Charles J. (1991) Constrained Maximum Likelihood in Logistic

However, do you want to...

- non-linear transformation of link often unappealing, distorts shape!
- Wand (2011) uses (constrained) spline to fit binary choice

Multivariate shapes

- rather than additive (cf Stout 2011)

Testing theories based on shapes

Design: no less important here than in RCM

- case selection

  minimizing confounders

  Eg., theories of campaign finance and “open seat” races
- selection of a test / distance-metric

  identifying unique and invariant implications from theory

  E.g., agenda theories hinge on status quo locations of (potential) proposals
- sensitivity analysis: bounds from theory and data

  E.g., what (implausible) distribution of SQ could make agenda theories observationally equivalent