Smoothing sparsely observed discrete distributions

P. E. Oliveira

Introduction

A “real world” motivation

- Forensic medicine, anthropological studies
- Given a corpse, characterize the age of death
  - probabilistic characterization
  - $\mathbb{P}$(age of death $\in$ interval) = level of confidence
- measure some parameters (teeth, bones, etc.)
- training data: measured parameters for fully known corpses

A “real world” motivation

- Forensic medicine, anthropological studies
- Given a corpse, characterize the age of death
  - probabilistic characterization
  - P(age of death $\in$ interval) = level of confidence
- measure some parameters (teeth, bones, etc.)

probability distribution for age of death

- training data: measured parameters for fully known corpses

A “real world” motivation

- Forensic medicine, anthropological studies
- Given a corpse, characterize the age of death
  - probabilistic characterization
  - $\mathbb{P}$(age of death $\in$ interval) = level of confidence
- measure some parameters (teeth, bones, etc.)

probability distribution for age of death

- training data: measured parameters for fully known corpses

The original problem - maths

- X vector of parameters: high dimensional, usually categorical
- Y age: integer valued, usually categorized
- estimate Y given $X = x$,
- First approaches: linear or multilinear regression
  numerical estimate, no probabilistic characterization

  Gustafon (1950), Konigsberg, Frankenberg, Walker (1994)
- alternatively, approximate $\mathrm{P}_{Y|X=x}$
- late 1990's: statisticians came into play
  use training data to estimate $\mathrm{P}_{(X,Y)}$ and P_Y
  Bayes Theorem to find $\mathrm{P}_{Y|X=x}$

  Lucy, Aykroyd, Pollard (1996), Lucy, Aykroyd, Pollard, Solheim (2002)
- what if P_Y is known?

The original problem - maths

- X vector of parameters: high dimensional, usually categorical
- Y age: integer valued, usually categorized
- estimate Y given $X = x$,
- First approaches: linear or multilinear regression
  numerical estimate, no probabilistic characterization
  Gustafon (1950), Konigsberg, Frankenberg, Walker (1994)
- alternatively, approximate $\mathrm{P}_{Y|X=x}$
- late 1990's: statisticians came into play
  use training data to estimate $\mathrm{P}_{(X,Y)}$ and P_Y
  Bayes Theorem to find $\mathrm{P}_{Y|X=x}$
  Lucy, Aykroyd, Pollard (1996), Lucy, Aykroyd, Pollard, Solheim (2002)
- what if P_Y is known?

The original problem - maths

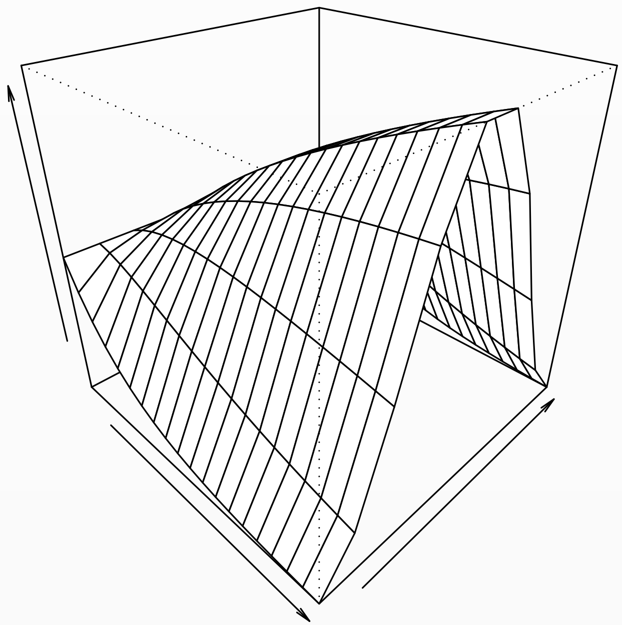
- X vector of parameters: high dimensional, usually categorical
- Y age: integer valued, usually categorized
- estimate Y given $X = x$,
- First approaches: linear or multilinear regression
  numerical estimate, no probabilistic characterization
  Gustafon (1950), Konigsberg, Frankenberg, Walker (1994)
- alternatively, approximate $\mathrm{P}_{Y|X=x}$
- late 1990's: statisticians came into play
  use training data to estimate $\mathrm{P}_{(X,Y)}$ and P_Y
  Bayes Theorem to find $\mathrm{P}_{Y|X=x}$
  Lucy, Aykroyd, Pollard (1996), Lucy, Aykroyd, Pollard, Solheim (2002)
- what if P_Y is known?

The original problem - maths

- X vector of parameters: high dimensional, usually categorical
- Y age: integer valued, usually categorized
- estimate Y given $X = x$,
- First approaches: linear or multilinear regression
  numerical estimate, no probabilistic characterization
  Gustafon (1950), Konigsberg, Frankenberg, Walker (1994)
- alternatively, approximate $\mathrm{P}_{Y|X=x}$
- late 1990's: statisticians came into play
  use training data to estimate $\mathrm{P}_{(X,Y)}$ and P_Y
  Bayes Theorem to find $\mathrm{P}_{Y|X=x}$
  Lucy, Aykroyd, Pollard (1996), Lucy, Aykroyd, Pollard, Solheim (2002)
- what if P_Y is known?

Example

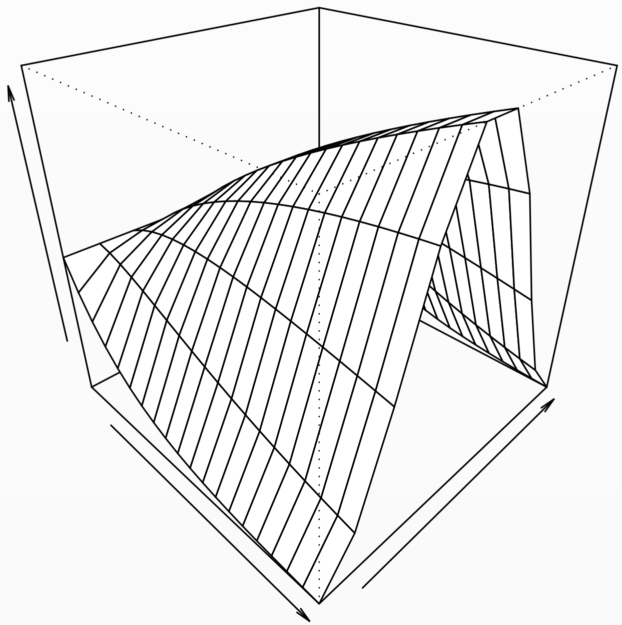
- 20×10 table; 150 points



0	1	1	0	2	1	1	2	0	1
0	0	0	2	0	0	2	0	0	0
0	0	0	0	1	0	0	0	2	1
0	1	0	0	1	0	2	0	1	1
0	0	0	0	1	0	1	1	2	0
2	1	2	2	1	1	0	2	0	0
0	2	1	1	1	2	0	1	1	0
0	1	0	0	1	4	0	0	1	0
0	1	1	1	1	1	1	2	2	0
0	0	2	2	1	1	2	0	1	0
0	0	0	3	0	2	1	0	1	0
0	0	1	1	1	1	3	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	2	0	2	0	0	1	0
0	1	0	2	4	4	2	2	1	0
0	0	0	3	1	3	1	0	1	0
0	0	0	0	4	0	1	0	0	0
0	0	0	0	4	1	1	0	0	0
0	0	0	1	2	0	0	4	1	0
0	1	1	2	2	2	0	1	0	0

Example

- 20×10 table; 150 points



0	1	1	0	2	1	1	2	0	1
0	0	0	2	0	0	2	0	0	0
0	0	0	0	1	0	0	0	2	1
0	1	0	0	1	0	2	0	1	1
0	0	0	0	1	0	1	1	2	0
2	1	2	2	1	1	0	2	0	0
0	2	1	1	1	2	0	1	1	0
0	1	0	0	1	4	0	0	1	0
0	1	1	1	1	1	1	2	2	0
0	0	2	2	1	1	2	0	1	0
0	0	0	3	0	2	1	0	1	0
0	0	1	1	1	1	3	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	2	0	2	0	0	1	0
0	1	0	2	4	4	2	2	1	0
0	0	0	3	1	3	1	0	1	0
0	0	0	0	4	0	1	0	0	0
0	0	0	0	4	1	1	0	0	0
0	0	0	1	2	0	0	4	1	0
0	1	1	2	2	2	0	1	0	0

- no more observations available; **sparse data**

Framework

The maths problem

- discrete distribution (1D, 2D)
  cells: $C_1, \ldots, C_k$ or $C_{i,j}$
  total number of cells: N
  cell probabilities (unknown): P_i or $P_{i,j}$
- some (typically few) observations: n
  cell counts: N_i or $N_{i,j}$
  cell frequencies: $\overline{P}_i = \frac{N_i}{n}$ or $\overline{P}_{i,j} = \frac{N_{i,j}}{n}$
- approximate the probability distribution
- 2D case: knowledge of marginal distribution
- nonasymptotic results

The maths problem

- discrete distribution (1D, 2D)
  cells: $C_1, \ldots, C_k$ or $C_{i,j}$
  total number of cells: N
  cell probabilities (unknown): P_i or $P_{i,j}$
- some (typically few) observations: n
  cell counts: N_i or $N_{i,j}$
  cell frequencies: $\overline{P}_i = \frac{N_i}{n}$ or $\overline{P}_{i,j} = \frac{N_{i,j}}{n}$
- approximate the probability distribution
- 2D case: knowledge of marginal distribution
- nonasymptotic results

Smoothing

- histogram is clearly a poor answer

  to many zeros, does not integrate knowledge of marginal distribution (2D)

0	1	1	0	2	1	1	2	0	1
0	0	0	2	0	0	2	0	0	0
0	0	0	0	1	0	0	0	2	1
0	1	0	0	1	0	2	0	1	1
0	0	0	0	1	0	1	1	2	0
2	1	2	2	1	1	0	2	0	0
0	2	1	1	1	2	0	1	1	0
0	1	0	0	1	4	0	0	1	0
0	1	1	1	1	1	1	2	2	0
0	0	2	2	1	1	2	0	1	0
0	0	0	3	0	2	1	0	1	0
0	0	1	1	1	1	3	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	2	0	2	0	0	1	0
0	1	0	2	4	4	2	2	1	0
0	0	0	3	1	3	1	0	1	0
0	0	0	0	4	0	1	0	0	0
0	0	0	0	4	1	1	0	0	0
0	0	0	1	2	0	0	4	1	0
0	1	1	2	2	2	0	1	0	0

Smoothing

- histogram is clearly a poor answer

  to many zeros, does not integrate knowledge of marginal distribution (2D)

  it can be (well) justified asymptotically
- smoothing methods: kernels and local polynomials

  continuous data
- categorical data: contiguity
- smoothed estimates
  - biased
  - good **mean square error** convergence rates
- how to incorporate marginal knowledge into smoothing?

Smoothing

- histogram is clearly a poor answer

  to many zeros, does not integrate knowledge of marginal distribution (2D)

  it can be (well) justified asymptotically

- smoothing methods: kernels and local polynomials

  continuous data

- categorical data: contiguity
- smoothed estimates
  - biased
  - good **mean square error** convergence rates
- how to incorporate marginal knowledge into smoothing?

Smoothing

- histogram is clearly a poor answer

  to many zeros, does not integrate knowledge of marginal distribution (2D)

  it can be (well) justified asymptotically

- smoothing methods: kernels and local polynomials

  continuous data

- categorical data: contiguity
- smoothed estimates
  - biased
  - good **mean square error** convergence rates
- how to incorporate marginal knowledge into smoothing?

Smoothing

- histogram is clearly a poor answer

  to many zeros, does not integrate knowledge of marginal distribution (2D)

  it can be (well) justified asymptotically

- smoothing methods: kernels and local polynomials

  continuous data

- categorical data: contiguity
- smoothed estimates
  - biased
  - good **mean square error** convergence rates
- how to incorporate marginal knowledge into smoothing?

Smoothing discrete data

- two early approaches:
  - ”convolution” with a prescribed distribution
  - discretized version of kernel methods

Simonoff (1983), Titterington, Bowman (1985)

Hall, Titterington (1987), Burman (1987)

- discretized polynomial method

Simonoff (1995, 1996), Dong, Simonoff (1995)

Aerts, Augustyns, Janssen (1997)

- asymptotic characterizations keeping sparseness

$$\frac{n}{N} \longrightarrow \lambda \text{ finite}$$

number of cells increases with number of observations

Smoothing discrete data

- two early approaches:
  - ”convolution” with a prescribed distribution
  - discretized version of kernel methods

Simonoff (1983), Titterington, Bowman (1985)

Hall, Titterington (1987), Burman (1987)

- discretized polynomial method

Simonoff (1995, 1996), Dong, Simonoff (1995)

Aerts, Augustyns, Janssen (1997)

- asymptotic characterizations keeping sparseness

$$\frac{n}{N} \longrightarrow \lambda \text{ finite}$$

number of cells increases with number of observations

Smoothing discrete data

- two early approaches:
  - ”convolution” with a prescribed distribution
  - discretized version of kernel methods

Simonoff (1983), Titterington, Bowman (1985)

Hall, Titterington (1987), Burman (1987)

- discretized polynomial method

Simonoff (1995, 1996), Dong, Simonoff (1995)

Aerts, Augustyns, Janssen (1997)

- asymptotic characterizations keeping sparseness

$$\frac{n}{N} \longrightarrow \lambda \text{ finite}$$

number of cells increases with number of observations

The estimators

poly. smoother 1D

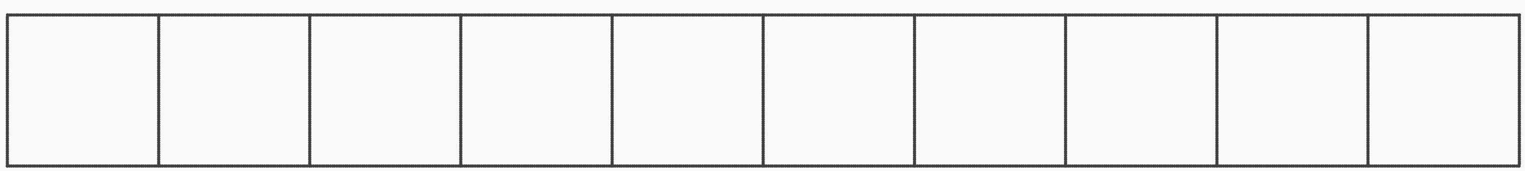
- symmetric weight function: $p(j),\ j = -u, \dots, u$
- $H_\ell = \sum_{i=1}^{N} \left(\overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p \right)^2 p(i - \ell)$
- $\mathbf{X}_\ell = \begin{bmatrix} & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{bmatrix}$

  $\mathbf{W}_\ell = \mathrm{diag}\Big(p(j - \ell),\ j = 1 - k, \dots, 2k \Big)$
- minimize $H_\ell = (\overline{\mathbf{P}} - \mathbf{X}_\ell \beta_\ell)^t \mathbf{W}_\ell (\overline{\mathbf{P}} - \mathbf{X}_\ell \beta_\ell)$
- The border effect

poly. smoother 1D

- symmetric weight function: $p(j),\ j = -u, \ldots, u$
- $$H_\ell = \sum_{i=1}^{N} \Big(\overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p\Big)^2 p(i-\ell)$$
- $$\mathbf{X}_\ell = \begin{bmatrix} & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{bmatrix}$$

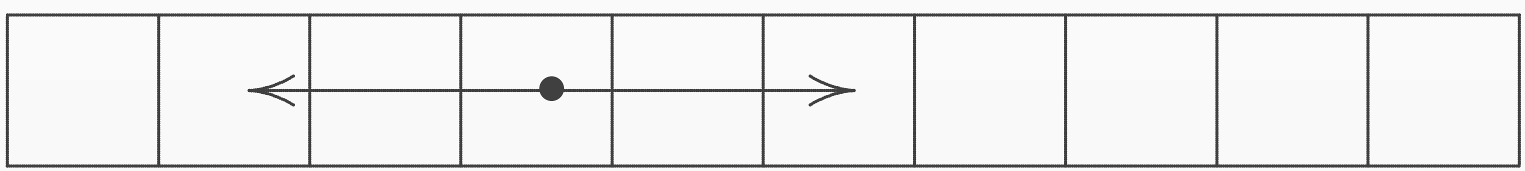
  $$\mathbf{W}_\ell = \mathrm{diag}\Big(p(j-\ell),\ j = 1-k, \ldots, 2k\Big)$$
- minimize $H_\ell = (\overline{\mathbf{P}} - \mathbf{X}_\ell\beta_\ell)^t \mathbf{W}_\ell (\overline{\mathbf{P}} - \mathbf{X}_\ell\beta_\ell)$
- The border effect



poly. smoother 1D

- symmetric weight function: $p(j),\ j=-u,\dots,u$
- $H_\ell = \sum_{i=1}^{N} \Big(\overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p\Big)^2 p(i-\ell)$
- $$\mathbf{X}_\ell = \begin{bmatrix} & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{bmatrix}$$

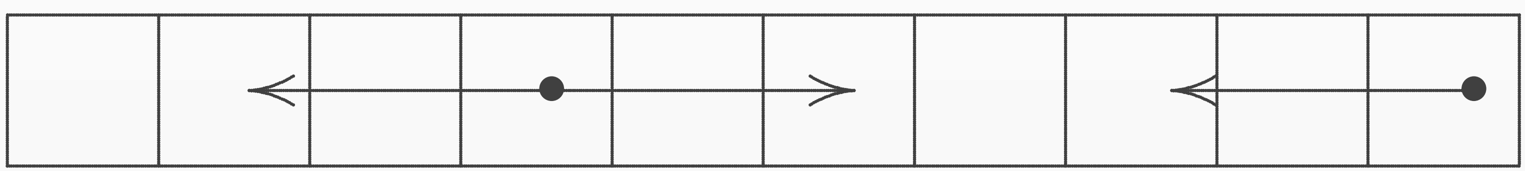
  $$\mathbf{W}_\ell = \operatorname{diag}\Big(p(j-\ell),\ j = 1-k,\dots,2k\Big)$$
- minimize $H_\ell = (\overline{\mathbf{P}} - \mathbf{X}_\ell \beta_\ell)^t \mathbf{W}_\ell (\overline{\mathbf{P}} - \mathbf{X}_\ell \beta_\ell)$
- The border effect



poly. smoother 1D

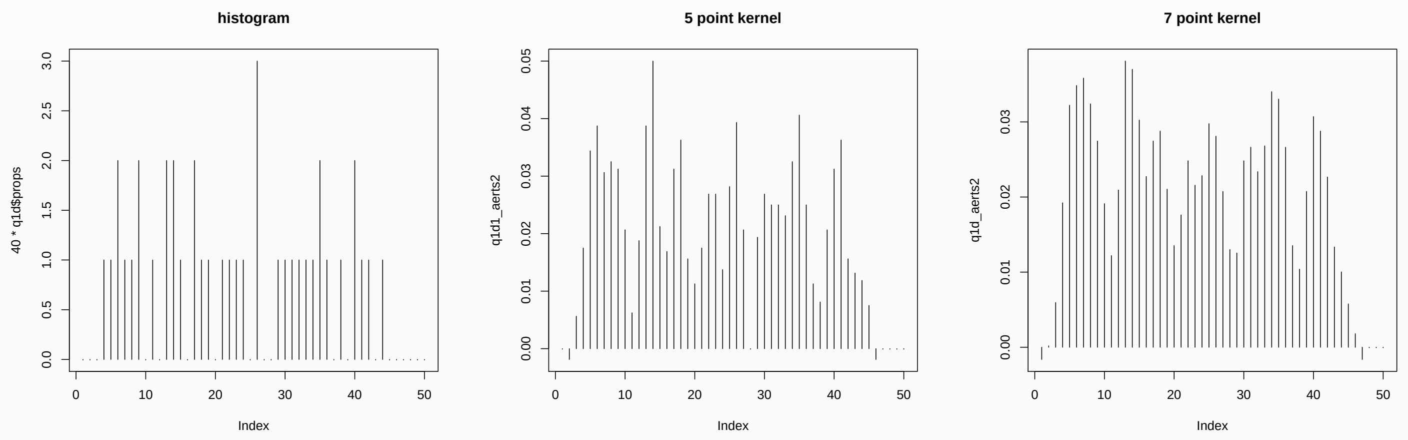
- symmetric weight function: $p(j),\ j=-u,\dots,u$
- $$H_\ell = \sum_{i=1}^{N}\Big(\overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p\Big)^2 p(i-\ell)$$
- $$\mathbf{X}_\ell = \begin{bmatrix} & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \\ & \cdots\cdots & \cdots\cdots & \cdots\cdots \end{bmatrix}$$

  $$\mathbf{W}_\ell = \operatorname{diag}\Big(p(j-\ell),\ j = 1-k,\dots,2k\Big)$$
- minimize $H_\ell = (\overline{\mathbf{P}} - \mathbf{X}_\ell\beta_\ell)^t\mathbf{W}_\ell(\overline{\mathbf{P}} - \mathbf{X}_\ell\beta_\ell)$
- The border effect



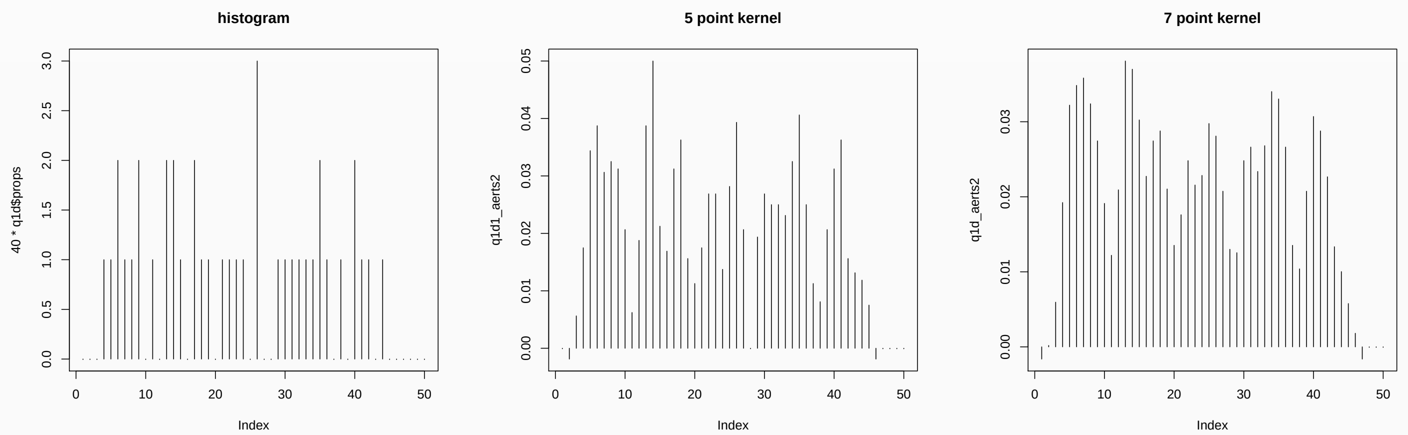
poly. smoother 1D

- $\mathrm{PS}_\ell(0) = \mathrm{PS}_\ell(1) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} p(j)$
- $\mathrm{PS}_\ell(2) = \mathrm{PS}_\ell(3) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} q(j)$ $\quad \sum_{j=-u}^{u} j^2 q(j) = 0$



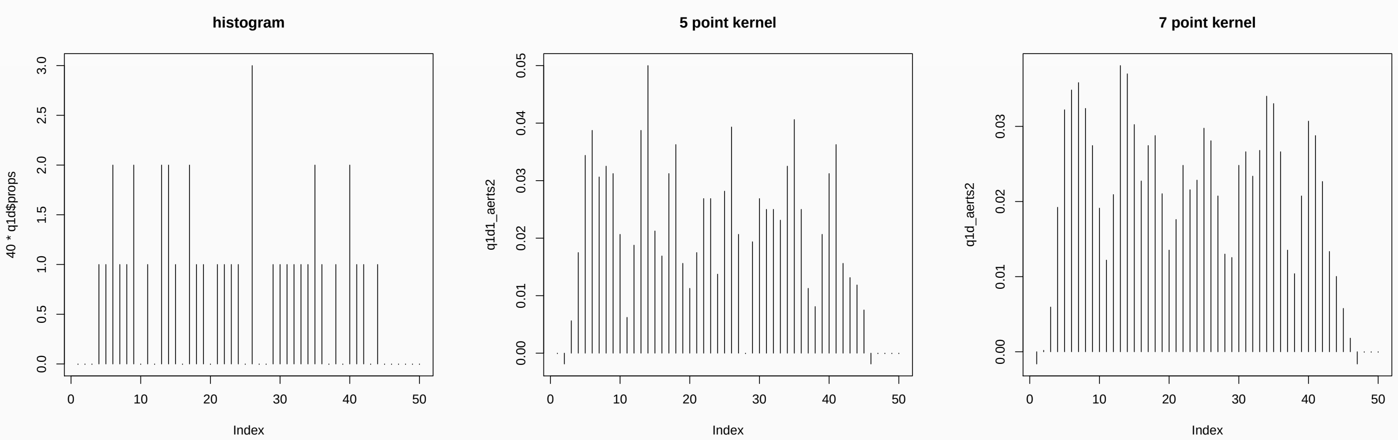
poly. smoother 1D

- $$\mathrm{PS}_\ell(0) = \mathrm{PS}_\ell(1) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} p(j)$$
- $$\mathrm{PS}_\ell(2) = \mathrm{PS}_\ell(3) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} q(j) \qquad \sum_{j=-u}^{u} j^2 q(j) = 0$$



poly. smoother 1D

- $$\mathrm{PS}_\ell(0) = \mathrm{PS}_\ell(1) = \sum_{j=-u}^{u} \overline{P}_{\ell-j}\, p(j)$$
- $$\mathrm{PS}_\ell(2) = \mathrm{PS}_\ell(3) = \sum_{j=-u}^{u} \overline{P}_{\ell-j}\, q(j) \qquad \sum_{j=-u}^{u} j^2 q(j) = 0$$



2D – degree 2

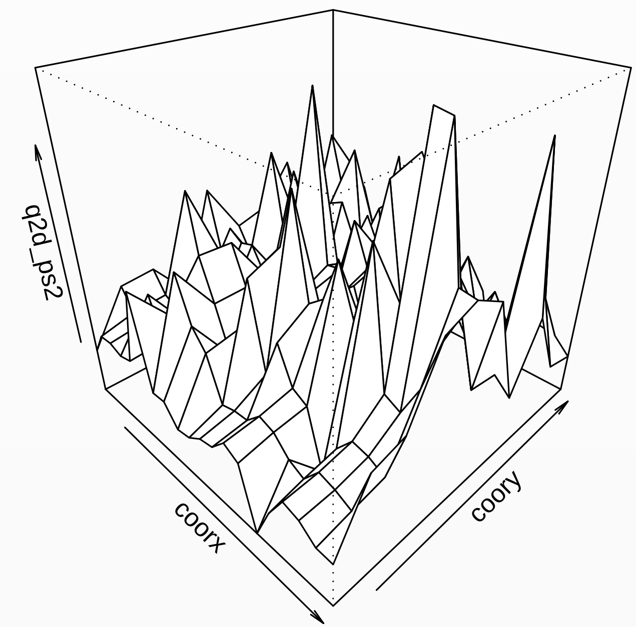
0	1	1	0	2	1	1	2	0	1
0	0	0	2	0	0	2	0	0	0
0	0	0	0	1	0	0	0	2	1
0	1	0	0	1	0	2	0	1	1
0	0	0	0	1	0	1	1	2	0
2	1	2	2	1	1	0	2	0	0
0	2	1	1	1	2	0	1	1	0
0	1	0	0	1	4	0	0	1	0
0	1	1	1	1	1	1	2	2	0
0	0	2	2	1	1	2	0	1	0
0	0	0	3	0	2	1	0	1	0
0	0	1	1	1	1	3	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	2	0	2	0	0	1	0
0	1	0	2	4	4	2	2	1	0
0	0	0	3	1	3	1	0	1	0
0	0	0	0	4	0	1	0	0	0
0	0	0	0	4	1	1	0	0	0
0	0	0	1	2	0	0	4	1	0
0	1	1	2	2	2	0	1	0	0

2D – degree 2

-9.965128e-04	4.721981e-03	0.0050759045	0.0008349213	0.011245185	0.005891897	0.0068352396	1.109993e-02	0.0001099982	5.181
1.580869e-03	2.505594e-03	0.0028480877	0.0129416120	0.004154047	0.003234309	0.0137405439	3.428898e-03	0.0028558472	2.710
1.710888e-03	2.210609e-03	0.0020490101	0.0028498377	0.008202191	0.002364985	0.0039309906	3.153937e-03	0.0142188067	9.308
1.195192e-03	6.179673e-03	0.0012283798	0.0015325213	0.007227906	0.001729657	0.0126851933	2.428458e-03	0.0083725779	7.420
1.603390e-03	1.884026e-03	0.0019010164	0.0021200014	0.007657558	0.002408935	0.0080221977	8.466088e-03	0.0134219006	2.514
9.754467e-03	5.609060e-03	0.0101415229	0.0100148385	0.005038259	0.004882510	-0.0010643480	9.150874e-03	-0.0009443339	-2.582
2.854394e-05	1.078408e-02	0.0058994051	0.0059958295	0.006100070	0.012021082	-0.0001077673	5.560378e-03	0.0050021814	-1.283
1.255469e-04	6.190997e-03	0.0011972852	0.0017938263	0.007530215	0.023647733	0.0018353259	1.680412e-03	0.0061719271	-1.732
-1.897867e-03	4.262225e-03	0.0049243643	0.0055297966	0.005198812	0.006755134	0.0056425323	1.021427e-02	0.0105918161	-1.221
-1.630505e-03	-3.553853e-04	0.0108517022	0.0122998861	0.006265007	0.007071982	0.0113364697	2.117729e-04	0.0052805899	-1.331
-5.586377e-04	9.281415e-05	0.0018946391	0.0181717885	0.002662193	0.012745567	0.0081851784	9.241139e-04	0.0059082695	-2.592
-2.289896e-04	1.929379e-04	0.0062052527	0.0082888318	0.007344061	0.008572015	0.0181968990	1.015544e-03	0.0007384092	-3.249
1.453278e-03	1.776886e-03	0.0024555204	0.0095358034	0.009180380	0.010124910	0.0092945726	2.406983e-03	0.0022849033	1.486
5.669475e-04	1.230106e-03	0.0015932888	0.0139858971	0.004451404	0.014817197	0.0033674921	2.118594e-03	0.0067610200	1.108
-5.909744e-03	-4.582776e-04	-0.0045938744	0.0091649089	0.019997531	0.021075494	0.0086148683	6.670175e-03	0.0009943413	-5.555
-2.063321e-03	-1.272290e-03	-0.0006801415	0.0169188400	0.009674483	0.018249365	0.0064487073	-2.864721e-05	0.0042748223	-1.521
6.311215e-04	7.933708e-04	0.0013509305	0.0038614380	0.026280996	0.004806888	0.0081023080	2.002920e-03	0.0014955388	6.744
1.492060e-04	3.035544e-04	0.0009891848	0.0030456233	0.025979395	0.008602320	0.0074486578	2.353181e-03	0.0009141857	2.146
-1.166191e-03	-5.051119e-04	0.0001104707	0.0064557935	0.013695496	0.001907090	0.0013933920	2.227397e-02	0.0059466016	-1.115
-1.268241e-03	4.747879e-03	0.0055811593	0.0123048729	0.013432385	0.011589534	-0.0001675233	6.449686e-03	-0.0009287611	-1.740

2D – degree 2

0	1	1	0	2	1	1	2	0	1
0	0	0	2	0	0	2	0	0	0
0	0	0	0	1	0	0	0	2	1
0	1	0	0	1	0	2	0	1	1
0	0	0	0	1	0	1	1	2	0
2	1	2	2	1	1	0	2	0	0
0	2	1	1	1	2	0	1	1	0
0	1	0	0	1	4	0	0	1	0
0	1	1	1	1	1	1	2	2	0
0	0	2	2	1	1	2	0	1	0
0	0	0	3	0	2	1	0	1	0
0	0	1	1	1	1	3	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	2	0	2	0	0	1	0
0	1	0	2	4	4	2	2	1	0
0	0	0	3	1	3	1	0	1	0
0	0	0	0	4	0	1	0	0	0
0	0	0	0	4	1	1	0	0	0
0	0	0	1	2	0	0	4	1	0
0	1	1	2	2	2	0	1	0	0



$$\mathrm{PS}_{i,j}(2) = \sum_{s,t} R_{s,t} \overline{P}_{s,t}$$

$$R_{s,t} = p_1(s-i)p_2(t-j)\left[U + V\left(\frac{s-i}{K}\right)^2 + W\left(\frac{t-j}{L}\right)^2\right]$$

2D with marginal

- $$H_{i,j} = \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j}\right)^t \mathbf{W}_{i,j} \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j}\right)$$
- $$\text{minimize} \sum_{\ell=1}^{L} H_{i,\ell}$$
  $$\text{subject to } \textstyle\sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i$$
- $$\mathrm{CPS}_{i,j}(p) = \mathrm{PS}_{i,j}(p) + \frac{1}{L}\left(\Pi_i - \sum_{\ell=1}^{L} \mathrm{PS}_{i,\ell}(p)\right)$$

- the constraint introduces an additive correction
- it does not prevent negative estimates
- corrects deviations from prescribed marginal uniformly along each line

2D with marginal

- $$H_{i,j} = \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j}\right)^t \mathbf{W}_{i,j} \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j}\right)$$
- $$\text{minimize} \sum_{\ell=1}^{L} H_{i,\ell}$$
  $$\text{subject to } \textstyle\sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i$$
- $$\mathrm{CPS}_{i,j}(p) = \mathrm{PS}_{i,j}(p) + \frac{1}{L}\left(\Pi_i - \sum_{\ell=1}^{L} \mathrm{PS}_{i,\ell}(p)\right)$$

- the constraint introduces an additive correction
- it does not prevent negative estimates
- corrects deviations from prescribed marginal uniformly along each line

Smoothing and nonnegativity

- Relative errors

$$H^*_{i,j} = \frac{1}{\beta_{0,i,j}} \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j} \right)^t \mathbf{W}_{i,j} \left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,j}\beta_{i,j} \right)$$

- χ_2 tests error

- $$\text{minimize } H^*_i = \sum_{\ell=1}^{L} H^*_{i,\ell}$$

$$\text{subject to } \sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i$$

one optimization problem per line

- 1D version needs a constraint: $\displaystyle\sum_{\ell=1}^{k} \beta_{0,\ell} = 1$

Solving the estimators

Some maths

- First order conditions $\qquad \mathbf{h} = (1,0,0,0,0,0)^t$

$$\lambda\beta_{0,i,\ell}\mathbf{h} =$$
$$= -2\mathbf{X}_{i,\ell}^t\mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - X_{i,\ell}\beta_{i,\ell}\right)$$
$$-\frac{\mathbf{h}}{\beta_{0,i,\ell}}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)^t\mathbf{W}_{i,j}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)$$

- $\mathbf{X}_{i,j} = [\mathbf{e}^t|\mathbf{T}] \longrightarrow \widetilde{\mathbf{X}}_{i,j} = [\mathbf{0}|\mathbf{T}] \qquad \widetilde{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell}\mathbf{h}$
- $\widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right) = \widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\widetilde{\mathbf{X}}_{i,\ell}\widetilde{\beta}_{i,\ell}$

$$\begin{bmatrix} 0 & \cdots\ 0 \\ \vdots & \\ 0 & \mathbf{M} \end{bmatrix}$$

Some maths

- First order conditions $\mathbf{h} = (1,0,0,0,0,0)^t$

$$\lambda\beta_{0,i,\ell}\mathbf{h} =$$

$$= -2\mathbf{X}_{i,\ell}^t \mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - X_{i,\ell}\beta_{i,\ell}\right)$$

$$-\frac{\mathbf{h}}{\beta_{0,i,\ell}}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)^t \mathbf{W}_{i,j}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)$$

- $\mathbf{X}_{i,j} = [\mathbf{e}^t|\mathbf{T}] \longrightarrow \widetilde{\mathbf{X}}_{i,j} = [\mathbf{0}|\mathbf{T}]$ $\qquad \widetilde{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell}\mathbf{h}$
- $\widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right) = \widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell}\widetilde{\mathbf{X}}_{i,\ell}\widetilde{\beta}_{i,\ell}$

$$\begin{bmatrix} 0 & \cdots\ 0 \\ \vdots & \mathbf{M} \\ 0 & \end{bmatrix}^{-1} = \begin{bmatrix} 0 & \cdots\ 0 \\ \vdots & \mathbf{M}^{-1} \\ 0 & \end{bmatrix}$$

Some maths

- First order conditions $\qquad \mathbf{h} = (1,0,0,0,0,0)^t$

$$\lambda\beta_{0,i,\ell}\mathbf{h} =$$
$$= -2\mathbf{X}_{i,\ell}^t\mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - X_{i,\ell}\beta_{i,\ell}\right)$$
$$-\frac{\mathbf{h}}{\beta_{0,i,\ell}}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)^t\mathbf{W}_{i,j}\left(\overrightarrow{\mathbf{P}} - \mathbf{X}_{i,\ell}\beta_{i,\ell}\right)$$

- $\mathbf{X}_{i,j} = [\mathbf{e}^t|\mathbf{T}] \longrightarrow \widetilde{\mathbf{X}}_{i,j} = [\mathbf{0}|\mathbf{T}] \qquad \widetilde{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell}\mathbf{h}$
- $\widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right) = \widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\widetilde{\mathbf{X}}_{i,\ell}\widetilde{\beta}_{i,\ell}$

$$\widetilde{\beta}_{i,\ell} = \left(\widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\widetilde{\mathbf{X}}_{i,\ell}\right)^{-1}\widetilde{\mathbf{X}}_{i,j}^t\mathbf{W}_{i,\ell}\left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right)$$

Some more maths

- $\widetilde{\beta}_{i,\ell} = \left(\widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell} \widetilde{\mathbf{X}}_{i,\ell}\right)^{-1} \widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell} \left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right)$
- reintroduce into the first (nonlinear) equation

  some matrix algebra

$$\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}^2 \mathbf{e}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \mathbf{e} = \lambda \beta_{0,i,\ell}^2$$

  and, using the marginal constraint,

$$\mathrm{CRPS}_{i,j} = \Pi_i \frac{\left|\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \overrightarrow{\mathbf{P}}\right|^{1/2}}{\sum_{\ell=1}^{L} \left|\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \overrightarrow{\mathbf{P}}\right|^{1/2}}$$

$$\mathbf{A}_{i,j} = \mathbf{W}_{i,j} \widetilde{\mathbf{X}}_{i,j} \left(\widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,j} \widetilde{\mathbf{X}}_{i,j}\right)^{-1} \widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,j}$$

- multiplicative correction to unconstrained solution

Some more maths

- $\widetilde{\beta}_{i,\ell} = \left(\widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell} \widetilde{\mathbf{X}}_{i,\ell}\right)^{-1} \widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,\ell} \left(\overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}\mathbf{e}\right)$
- reintroduce into the first (nonlinear) equation

  some matrix algebra

  $$\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \overrightarrow{\mathbf{P}} - \beta_{0,i,\ell}^2 \mathbf{e}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \mathbf{e} = \lambda \beta_{0,i,\ell}^2$$

  and, using the marginal constraint,

  $$\mathrm{CRPS}_{i,j} = \Pi_i \frac{\left|\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \overrightarrow{\mathbf{P}}\right|^{1/2}}{\sum_{\ell=1}^{L} \left|\overrightarrow{\mathbf{P}}^t \left(\mathbf{W}_{i,\ell} - \mathbf{A}_{i,\ell}\right) \overrightarrow{\mathbf{P}}\right|^{1/2}}$$

  $$\mathbf{A}_{i,j} = \mathbf{W}_{i,j} \widetilde{\mathbf{X}}_{i,j} \left(\widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,j} \widetilde{\mathbf{X}}_{i,j}\right)^{-1} \widetilde{\mathbf{X}}_{i,j}^t \mathbf{W}_{i,j}$$

- multiplicative correction to unconstrained solution

Even more maths

- $$\mathbf{e}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \mathbf{e} = 1 - \left(\frac{\tau_2^4\sigma_1^4 + \tau_1^4\sigma_2^4 + 2\sigma_1^4\sigma_2^4}{\delta}\right)$$

$$\delta = \tau_1^4\tau_2^4 - \sigma_1^4\sigma_2^4$$

- $\mathbf{e}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \mathbf{e} = 0$ for “gaussian kernels”

$$\tau_1^4 - 3\sigma_1^4 = 0, \quad \tau_2^4 - 3\sigma_2^4 = 0$$

- choose leptokurtic kernels

- some cheating – border and edge effects versus replication of data

Even more maths

- $$\mathbf{e}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \mathbf{e} = 1 - \left(\frac{\tau_2^4 \sigma_1^4 + \tau_1^4 \sigma_2^4 + 2\sigma_1^4 \sigma_2^4}{\delta} \right)$$

$$\delta = \tau_1^4 \tau_2^4 - \sigma_1^4 \sigma_2^4$$

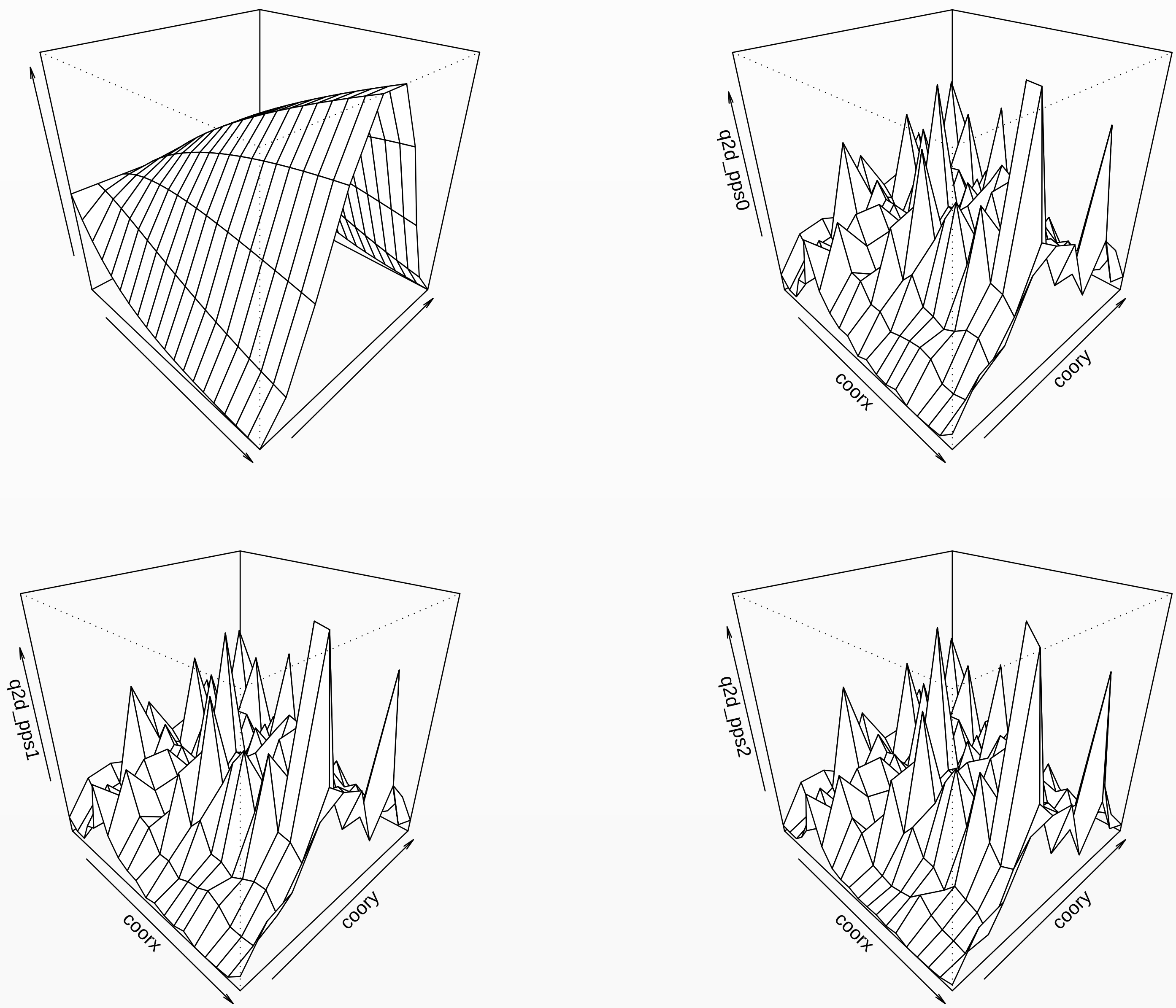
- $\mathbf{e}^t \left(\mathbf{W}_{i,j} - \mathbf{A}_{i,j}\right) \mathbf{e} = 0$ for "gaussian kernels"

$$\tau_1^4 - 3\sigma_1^4 = 0, \quad \tau_2^4 - 3\sigma_2^4 = 0$$

- choose leptokurtic kernels
- some cheating – border and edge effects versus replication of data

Simulations

2D examples


q2d_pps0
coorx
coory
q2d_pps1
coorx
coory
q2d_pps2
coorx
coory

comparing

- minimum estimates

CPS(2)	-0.005909744	CRPS(2)	0.0003165518
CRPS(1)	0.0003303732	CRPS(2)	0.0003733672
true dist. $\longrightarrow$ 0.0001364			

- error criteria
  - $$\mathrm{MSSE}(\mathbf{P}^*) = \mathrm{E}\left(\sum_{i=1}^{k}(P_i^* - P_i)^2\right)$$
  - $$\mathrm{NSUP}(\mathbf{P}^*) = \sup_{1\le i\le k} |P_i^* - P_i|$$
  - $$\mathrm{SPSUP}(\mathbf{P}^*) = \sup_{1\le i\le k} \left|\frac{P_i^*}{P_i} - 1\right|$$ Simonof (1996)

2D versions: adapt summations

2D – comparing

From this one estimate:

- MSSE

CPS(2)	0.004213980	CRPS(0)	0.001135212
CRPS(1)	0.001260446	CRPS(2)	0.002095495

- NSUP

CPS(2)	0.01753402	CRPS(0)	0.009084917
CRPS(1)	0.009717237	CRPS(2)	0.01282403

- SPSUP

CPS(2)	18.61315	CRPS(0)	19.30157
CRPS(1)	17.59961	CRPS(2)	9.2398

Comparing – Monte Carlo

MSSE

CPS(0)	0.002455097	CRPS(0)	0.001103970
CPS(2)	0.004304129	CRPS(1)	0.001194746
		CRPS(2)	0.002019461