Smoothing sparsely observed discrete distributions

P. E. Oliveira
Introduction

- Motivation
- Formulation
- Example

Framework

The estimators

Solving the estimators

Simulations
A “real world” motivation

- Forensic medicine, anthropological studies
  - Given a corpse, characterize the age of death
    - probabilistic characterization
    - \( P(\text{age of death} \in \text{interval}) = \text{level of confidence} \)
  - 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
  - $P(\text{age of death } \in \text{ interval}) = \text{ level of confidence}$
- 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
  - \( P(\text{age of death} \in \text{interval}) = \text{level of confidence} \)
- measure some parameters (teeth, bones, etc.)
- 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 $P_{Y|X=x}$
- late 1990’s: statisticians came into play
  use training data to estimate $P_{(X,Y)}$ and $P_Y$
  Bayes Theorem to find $P_{Y|X=x}$
- 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 $P_{Y|X=x}$
- late 1990’s: statisticians came into play
  use training data to estimate $P_{(X,Y)}$ and $P_Y$
  Bayes Theorem to find $P_{Y|X=x}$
- 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 $P_{Y|X=x}$
  
  late 1990's: statisticians came into play
  use training data to estimate $P_{(X,Y)}$ and $P_Y$
  Bayes Theorem to find $P_{Y|X=x}$
  
  
  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 \( P_{Y|X=x} \)
- late 1990’s: statisticians came into play
  use training data to estimate \( P_{(X,Y)} \) and \( P_Y \)
  Bayes Theorem to find \( P_{Y|X=x} \)
- what if \( P_Y \) is known?
Example

- \(20 \times 10\) table; 150 points
Example

- 20 × 10 table; 150 points

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

Framework
- The maths problem
- Smoothing
- Discrete data

The estimators

Solving the estimators

Simulations

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  0  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
  Aerts, Augustyns, Janssen (1997)

- asymptotic characterizations keeping sparseness
  \[ \frac{n}{N} \rightarrow \lambda \text{ finite} \]
  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
  - Aerts, Augustyns, Janssen (1997)
- asymptotic characterizations keeping sparseness
  \[ \frac{n}{N} \rightarrow \lambda \text{ finite} \]
  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
  Aerts, Augustyns, Janssen (1997)

- asymptotic characterizations keeping sparseness

\[ \frac{n}{N} \rightarrow \lambda \text{ finite} \]

number of cells increases with number of observations
The estimators

- poly. smoother 1D
- 2D – degree 2
- 2D with marginal
- Nonnegativity
**poly. smoother 1D**

- symmetric weight function: \( p(j), \quad j = -u, \ldots, u \)
- \( H_\ell = \sum_{i=1}^{N} \left( \overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p \right)^2 p(i - \ell) \)
- \( X_\ell = \begin{bmatrix} 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \\ \vdots & \ddots & \ddots & \ddots \\ \vdots & \cdots & \ddots & \ddots \end{bmatrix} \)
- \( W_\ell = \text{diag}(p(j - \ell), \quad j = 1 - k, \ldots, 2k) \)
- minimize \( H_\ell = (\overline{P} - X_\ell \beta_\ell)^t W_\ell (\overline{P} - X_\ell \beta_\ell) \)
- The border effect
poly. smoother 1D

- symmetric weight function: \( p(j), j = -u, \ldots, u \)

- \( H_\ell = \sum_{i=1}^{N} \left( \overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p \right)^2 p(i-\ell) \)

- \( X_\ell = \begin{bmatrix} 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \end{bmatrix} \)

- \( W_\ell = \text{diag}\left( p(j-\ell), j = 1 - k, \ldots, 2k \right) \)

- minimize \( H_\ell = (\overline{P} - X_\ell \beta_\ell)^t W_\ell (\overline{P} - X_\ell \beta_\ell) \)

- The border effect
poly. smoother 1D

- symmetric weight function: \( p(j), j = -u, \ldots, u \)

\[
H_\ell = \sum_{i=1}^{N} \left( \overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p \right)^2 p(i - \ell)
\]

- \( X_\ell = \begin{bmatrix}
1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}
\]

- \( W_\ell = \text{diag}(p(j - \ell), j = 1 - k, \ldots, 2k) \)

- minimize \( H_\ell = (\overline{P} - X_\ell \beta_\ell)^t W_\ell (\overline{P} - X_\ell \beta_\ell) \)

- The border effect
poly. smoother 1D

- symmetric weight function: \( p(j), j = -u, \ldots, u \)
- \( H_\ell = \sum_{i=1}^{N} \left( \overline{P}_i - \beta_{0,\ell} - \beta_{1,\ell}(x_i - x_\ell) - \beta_{p,\ell}(x_i - x_\ell)^p \right)^2 p(i - \ell) \)
- \( X_\ell = \begin{bmatrix} 1 & x_i - x_\ell & \cdots & (x_i - x_\ell)^p \end{bmatrix} \)
- \( W_\ell = \text{diag} \left( p(j - \ell), j = 1 - k, \ldots, 2k \right) \)
- minimize \( H_\ell = (\overline{P} - X_\ell \beta_\ell)^t W_\ell (\overline{P} - X_\ell \beta_\ell) \)
- The border effect
poly. smoother 1D

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

\[
\begin{align*}
\text{histogram} & \quad \text{5 point kernel} & \quad \text{7 point kernel}
\end{align*}
\]
poly. smoother 1D

- $\text{PS}_\ell(0) = \text{PS}_\ell(1) = \sum_{j=-u}^{u} P_{\ell-j} p(j)$
- $\text{PS}_\ell(2) = \text{PS}_\ell(3) = \sum_{j=-u}^{u} P_{\ell-j} q(j)$
  \[ \sum_{j=-u}^{u} j^2 q(j) = 0 \]
poly. smoother 1D

- \( \text{PS}_\ell(0) = \text{PS}_\ell(1) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} p(j) \)

- \( \text{PS}_\ell(2) = \text{PS}_\ell(3) = \sum_{j=-u}^{u} \overline{P}_{\ell-j} q(j) \sum_{j=-u}^{u} j^2 q(j) = 0 \)

Solutions

- \( \text{PS}_\ell(0) \) and \( \text{PS}_\ell(1) \) are symmetric around 0.

- \( \text{PS}_\ell(2) \) and \( \text{PS}_\ell(3) \) are exploited to remove the non-negativity constraint.

Histograms

- 5 point kernel
- 7 point kernel
## 2D – degree 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
2D – degree 2

Introduction

Framework

The estimators

- poly. smoother 1D
- poly. smoother 1D
- 2D – degree 2
- 2D with marginal
- Nonnegativity

Solving the estimators

Simulations

Smoothing sparsely observed discrete distributions
2D – degree 2

\[
PS_{i,j}(2) = \sum_{s,t} R_{s,t} \overline{P}_{s,t}
\]

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

- \( H_{i,j} = \left( \overrightarrow{P} - X_{i,j} \beta_{i,j} \right)^t W_{i,j} \left( \overrightarrow{P} - X_{i,j} \beta_{i,j} \right) \)

- minimize \( \sum_{\ell=1}^{L} H_{i,\ell} \)
  subject to \( \sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i \)

- \( \text{CPS}_{i,j}(p) = \text{PS}_{i,j}(p) + \frac{1}{L} \left( \Pi_i - \sum_{\ell=1}^{L} \text{PS}_{i,\ell}(p) \right) \)

- the constraint introduces an additive correction
- it does not prevent negative estimates
- corrects deviations from prescribed marginal uniformly along each line
2D with marginal

- \( H_{i,j} = \left( \mathbf{P} - X_{i,j} \beta_{i,j} \right)^t W_{i,j} \left( \mathbf{P} - X_{i,j} \beta_{i,j} \right) \)

- minimize \( \sum_{\ell=1}^{L} H_{i,\ell} \)
  subject to \( \sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i \)

- \( \text{CPS}_{i,j}(p) = \text{PS}_{i,j}(p) + \frac{1}{L} \left( \Pi_i - \sum_{\ell=1}^{L} \text{PS}_{i,\ell}(p) \right) \)

- 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
  \[ H_{i,j}^* = \frac{1}{\beta_{0,i,j}} (\bar{P} - X_{i,j} \beta_{i,j})^t W_{i,j} (\bar{P} - X_{i,j} \beta_{i,j}) \]

- \( \chi_2 \) tests error

- minimize \( H_i^* = \sum_{\ell=1}^{L} H_{i,\ell}^* \)
  subject to \( \sum_{j=1}^{L} \beta_{0,i,j} = \Pi_i \)

- 1D version needs a constraint: \( \sum_{\ell=1}^{k} \beta_{0,\ell} = 1 \)
Solving the estimators

- Some maths
- Some more maths
- Even more maths

Simulations
Some maths

- First order conditions

\[ \lambda \beta_{0,i,\ell} h = \]
\[ = -2X_{i,\ell}^t W_{i,\ell} (\bar{P} - X_{i,\ell} \beta_{i,\ell}) \]
\[ - \frac{h}{\beta_{0,i,\ell}} (\bar{P} - X_{i,\ell} \beta_{i,\ell})^t W_{i,j} (\bar{P} - X_{i,\ell} \beta_{i,\ell}) \]

- \( X_{i,j} = [e^t | T] \) \( \longrightarrow \) \( \tilde{X}_{i,j} = [0 | T] \) \( \bar{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell} h \)

- \( \tilde{X}_{i,j}^t W_{i,\ell} (\bar{P} - \beta_{0,i,\ell} e) = \tilde{X}_{i,j}^t W_{i,\ell} \tilde{X}_{i,\ell} \bar{\beta}_{i,\ell} \)

\[
\begin{bmatrix}
0 & \cdots & 0 \\
\vdots & M \\
0 & \cdots & 0
\end{bmatrix}
\]
Some maths

- **First order conditions**

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

\[
= -2X_{i,\ell}^t W_{i,\ell} \left( \mathbf{P} - X_{i,\ell} \beta_{i,\ell} \right)
\]

\[
- \frac{\mathbf{h}}{\beta_{0,i,\ell}} \left( \mathbf{P} - X_{i,\ell} \beta_{i,\ell} \right)^t W_{i,j} \left( \mathbf{P} - X_{i,\ell} \beta_{i,\ell} \right)
\]

- \( \mathbf{X}_{i,j} = [e^{t}|T] \rightarrow \tilde{\mathbf{X}}_{i,j} = [0|T] \quad \tilde{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell} \mathbf{h} \)

- \( \tilde{\mathbf{X}}_{i,j}^t W_{i,\ell} \left( \mathbf{P} - \beta_{0,i,\ell} e \right) = \tilde{\mathbf{X}}_{i,j}^t W_{i,\ell} \tilde{\mathbf{X}}_{i,\ell} \tilde{\beta}_{i,\ell} \)

\[
\begin{bmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & M \\
\end{bmatrix}^{-1} = 
\begin{bmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & M^{-1} \\
\end{bmatrix}
\]
Some maths

- First order conditions

$$h = (1, 0, 0, 0, 0, 0)^t$$

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

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

$$- \frac{h}{\beta_{0,i,\ell}} \left( \overrightarrow{P} - X_{i,\ell} \beta_{i,\ell} \right)^t W_{i,j} \left( \overrightarrow{P} - X_{i,\ell} \beta_{i,\ell} \right)$$

- $$X_{i,j} = [e^t | T] \rightarrow \tilde{X}_{i,j} = [0 | T] \quad \tilde{\beta}_{i,\ell} = \beta_{i,\ell} - \beta_{0,i,\ell} h$$

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

$$\tilde{\beta}_{i,\ell} = \left( \tilde{X}_{i,j}^t W_{i,\ell} \tilde{X}_{i,\ell} \right)^{-1} \tilde{X}_{i,j}^t W_{i,\ell} \left( \overrightarrow{P} - \beta_{0,i,\ell} e \right)$$
Some more maths

- \( \tilde{\beta}_{i,\ell} = \left( \tilde{X}_{i,j}^t W_{i,\ell} \tilde{X}_{i,\ell} \right)^{-1} \tilde{X}_{i,j}^t W_{i,\ell} \left( \vec{P} - \beta_{0,i,\ell} e \right) \)

- reintroduce into the first (nonlinear) equation some matrix algebra

\[ \vec{P}^t (W_{i,\ell} - A_{i,\ell}) \vec{P} - \beta_{0,i,\ell}^2 e^t (W_{i,\ell} - A_{i,\ell}) e = \lambda \beta_{0,i,\ell}^2 \]

and, using the marginal constraint,

\[ CRPS_{i,j} = \prod_i \frac{\left| \vec{P}^t (W_{i,j} - A_{i,j}) \vec{P} \right|^{1/2}}{\sum_{\ell=1}^L \left| \vec{P}^t (W_{i,\ell} - A_{i,\ell}) \vec{P} \right|^{1/2}} \]

\[ A_{i,j} = W_{i,j} \tilde{X}_{i,j} \left( \tilde{X}_{i,j}^t W_{i,j} \tilde{X}_{i,j} \right)^{-1} \tilde{X}_{i,j}^t W_{i,j} \]

- multiplicative correction to unconstrained solution
Some more maths

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

- reintroduce into the first (nonlinear) equation some matrix algebra

\[
\overrightarrow{P}^t (W_{i,\ell} - A_{i,\ell}) \overrightarrow{P} - \beta_{0,i,\ell}^2 e^t (W_{i,\ell} - A_{i,\ell}) e = \lambda \beta_{0,i,\ell}^2
\]

and, using the marginal constraint,

\[
\text{CRPS}_{i,j} = \prod_i \frac{\left| \overrightarrow{P}^t (W_{i,j} - A_{i,j}) \overrightarrow{P} \right|^{1/2}}{\sum_{\ell=1}^L \left| \overrightarrow{P}^t (W_{i,\ell} - A_{i,\ell}) \overrightarrow{P} \right|^{1/2}}
\]

- multiplicative correction to unconstrained solution

- \( A_{i,j} = W_{i,j} \tilde{X}_{i,j} \left( \tilde{X}_{i,j}^t W_{i,j} \tilde{X}_{i,j} \right)^{-1} \tilde{X}_{i,j}^t W_{i,j} \)
**Even more maths**

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

\[ \delta = \tau_1^4 \tau_2^4 - \sigma_1^4 \sigma_2^4 \]

- $e^t (W_{i,j} - A_{i,j}) e = 0$ 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

\begin{itemize}
  \item \( e^t (W_{i,j} - A_{i,j}) e = 1 - \left( \frac{\tau_2^4 \sigma_1^4 + \tau_1^4 \sigma_2^4 + 2\sigma_1^4 \sigma_2^4}{\delta} \right) \)
  \end{itemize}

\[ \delta = \tau_1^4 \tau_2^4 - \sigma_1^4 \sigma_2^4 \]

\begin{itemize}
  \item \( e^t (W_{i,j} - A_{i,j}) e = 0 \) for “gaussian kernels”
  \end{itemize}

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

\begin{itemize}
  \item choose leptokurtic kernels
  \end{itemize}

\begin{itemize}
  \item some cheating – border and edge effects versus replication of data
  \end{itemize}
Simulations
2D examples
comparing

- minimum estimates

<table>
<thead>
<tr>
<th></th>
<th>CPS(2)</th>
<th>CRPS(2)</th>
<th>CRPS(1)</th>
<th>CRPS(2)</th>
<th>CRPS(2)</th>
<th>true dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS(2)</td>
<td>-0.005909744</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0001364</td>
</tr>
<tr>
<td>CRPS(1)</td>
<td>0.0003303732</td>
<td>0.0003165518</td>
<td>0.0003303732</td>
<td>0.0003733672</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- error criteria

  \[ \text{MSSE}(P^*) = E \left( \sum_{i=1}^{k} (P_i^* - P_i)^2 \right) \]

  \[ \text{NSUP}(P^*) = \sup_{1 \leq i \leq k} |P_i^* - P_i| \]

  \[ \text{SPSUP}(P^*) = \sup_{1 \leq i \leq k} \left| \frac{P_i^*}{P_i} - 1 \right| \]

Simonof (1996)

2D versions: adapt summations
2D – comparing

From this one estimate:

- **MSSE**

<table>
<thead>
<tr>
<th></th>
<th>CPS(2)</th>
<th>CRPS(0)</th>
<th>CRPS(1)</th>
<th>CRPS(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.004213980</td>
<td>0.001135212</td>
<td>0.001260446</td>
<td>0.002095495</td>
</tr>
</tbody>
</table>

- **NSUP**

<table>
<thead>
<tr>
<th></th>
<th>CPS(2)</th>
<th>CRPS(0)</th>
<th>CRPS(1)</th>
<th>CRPS(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01753402</td>
<td>0.009084917</td>
<td>0.009717237</td>
<td>0.01282403</td>
</tr>
</tbody>
</table>

- **SPSUP**

<table>
<thead>
<tr>
<th></th>
<th>CPS(2)</th>
<th>CRPS(0)</th>
<th>CRPS(1)</th>
<th>CRPS(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18.61315</td>
<td>19.30157</td>
<td>17.59961</td>
<td>9.2398</td>
</tr>
</tbody>
</table>
Comparing – Monte Carlo

Introduction
Framework
The estimators
Solving the estimators
Simulations
- 2D examples
- comparing
- 2D – comparing
- Comparing

<table>
<thead>
<tr>
<th></th>
<th>MSSE</th>
<th></th>
<th>NSUP</th>
<th></th>
<th>SPSUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS(0)</td>
<td>0.002455097</td>
<td>CRPS(0)</td>
<td>0.001103970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPS(2)</td>
<td>0.004304129</td>
<td>CRPS(1)</td>
<td>0.001194746</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CRPS(2)</td>
<td>0.002019461</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Smoothing sparsely observed discrete distributions