On the relationship between data sharpening and Firth's adjusted score function

John Braun & Patrick Brown

Data sharpening

Data sharpening

Two recipes - each data point is adjusted differently

Overview

Data sharpening

Two recipes - each data point is adjusted differently

Framework lacking

Overview

Data sharpening

Two recipes - each data point is adjusted differently

Framework lacking

Firth's adjusted score function

Overview

Data sharpening

Two recipes - each data point is adjusted differently

Framework lacking

Firth's adjusted score function

Derivative estimation - new recipes

Overview

Data sharpening

Two recipes - each data point is adjusted differently

Framework lacking

Firth's adjusted score function

Derivative estimation - new recipes

Unresolved extensions

Overview

Data sharpening

Two recipes - each data point is adjusted differently

Framework lacking

Firth's adjusted score function

Derivative estimation - new recipes

Unresolved extensions

Standard asymptotics - Wand & Jones (1995)

Data sharpening - Peter Hall

Data sharpening - Peter Hall

Preprocessing $\Longrightarrow$ standard software (KDE, N-W)

Data sharpening - Peter Hall

Preprocessing $\Longrightarrow$ standard software (KDE, N-W)

Reducing bias & enforcing constraints

Data sharpening - Peter Hall

Preprocessing $\Longrightarrow$ standard software (KDE, N-W)

Reducing bias & enforcing constraints

Density estimation & nonparametric regression

Data Sharpening

Data sharpening - Peter Hall

Preprocessing $\Longrightarrow$ standard software (KDE, N-W)

Reducing bias & enforcing constraints

Density estimation & nonparametric regression

Reduce bias without inflating variance

Data Sharpening

Data sharpening - Peter Hall

Preprocessing $\Longrightarrow$ standard software (KDE, N-W)

Reducing bias & enforcing constraints

Density estimation & nonparametric regression

Reduce bias without inflating variance

No reliance on data greedy alternatives

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad X_i \sim f$$

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad\qquad X_i \sim f$$

KDE: $\hat{f}(x) = \frac{1}{n} \sum K_h(X_i - x) \qquad\qquad \mathbf{E}[\hat{f}(x)] = f(x) + O(h^2)$

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad\qquad X_i \sim f$$

KDE: $\hat{f}(x) = \frac{1}{n}\sum K_h(X_i - x)$ $\qquad$ $\mathbf{E}[\hat{f}(x)] = f(x) + O(h^2)$

N-W: $\hat{g}(x) = \frac{\sum_i K_h(X_i - x)Y_i}{\sum K_h(X_i - x)}$ $\qquad$ $\mathbf{E}[\hat{g}(x)] = g(x) + O(h^2)$

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad\qquad X_i \sim f$$

KDE: $\hat{f}(x) = \frac{1}{n}\sum K_h(X_i - x)$ $\qquad$ $\mathbf{E}[\hat{f}(x)] = f(x) + O(h^2)$

N-W: $\hat{g}(x) = \frac{\sum_i K_h(X_i - x)\,Y_i}{\sum K_h(X_i - x)}$ $\qquad$ $\mathbf{E}[\hat{g}(x)] = g(x) + O(h^2)$

DS for N-W: $Y_i \longrightarrow Y_i^\star = Y_i + [\,Y_i - \hat{g}(X_i)]$

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad\qquad X_i \sim f$$

KDE: $\hat{f}(x) = \frac{1}{n}\sum K_h(X_i - x) \qquad \mathbf{E}[\hat{f}(x)] = f(x) + O(h^2)$

N-W: $\hat{g}(x) = \frac{\sum_i K_h(X_i - x)\, Y_i}{\sum K_h(X_i - x)} \qquad \mathbf{E}[\hat{g}(x)] = g(x) + O(h^2)$

DS for N-W: $Y_i \longrightarrow Y_i^\star = Y_i + [\, Y_i - \hat{g}(X_i)]$

DS for KDE: $X_i \longrightarrow X_i^\star = \frac{\sum_j K_h(X_j - X_i) X_j}{\sum_j K_h(X_j - X_i)}$

Data Sharpening

$$Y_i = g(X_i) + e_i \qquad\qquad X_i \sim f$$

KDE: $\hat{f}(x) = \frac{1}{n} \sum K_h(X_i - x) \qquad \mathbf{E}[\hat{f}(x)] = f(x) + O(h^2)$

N-W: $\hat{g}(x) = \frac{\sum_i K_h(X_i - x) Y_i}{\sum K_h(X_i - x)} \qquad \mathbf{E}[\hat{g}(x)] = g(x) + O(h^2)$

DS for N-W: $Y_i \longrightarrow Y_i^\star = Y_i + [\,Y_i - \hat{g}(X_i)]$

DS for KDE: $X_i \longrightarrow X_i^\star = \frac{\sum_j K_h(X_j - X_i) X_j}{\sum_j K_h(X_j - X_i)}$

DS for derivative estimation?

Density Estimation:

Density Estimation:

“Cluster design points near peaks”

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"

Data Sharpening

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

"...counteract bias from numerator and denominator"

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

"...counteract bias from numerator and denominator"
Intuition: Residuals capture bias

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

"...counteract bias from numerator and denominator"
Intuition: Residuals capture bias

Origins *unclear*

Data Sharpening

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

"...counteract bias from numerator and denominator"
Intuition: Residuals capture bias

Origins *unclear*

Framework lacking

Density Estimation:

"Cluster design points near peaks"
"Spread them further apart near troughs"
Some appeal to an asymptotic expression

Nonparametric regression

"...counteract bias from numerator and denominator"
Intuition: Residuals capture bias

Origins *unclear*

Framework lacking

Note: $\hat{f}(x), \quad \hat{g}(x)$ solve score equations

Firth's Adjusted Score Function

Parametric setting: $U(\theta) = \nabla \ell(\theta) \qquad \ell(\theta) = \log \mathcal{L}(\theta)$

Firth's Adjusted Score Function

Parametric setting: $U(\theta) = \nabla \ell(\theta) \qquad \ell(\theta) = \log \mathcal{L}(\theta)$

MLE: $\hat{\theta}$ **with bias** $b(\theta)$

Firth's Adjusted Score Function

Parametric setting: $U(\theta) = \nabla \ell(\theta) \qquad \ell(\theta) = \log \mathcal{L}(\theta)$

MLE: $\hat{\theta}$ **with bias** $b(\theta)$

Firth's idea: reduce bias of $\hat{\theta}$ **by adjusting** $U(\theta)$

$$U^{\star}(\theta) = U(\theta) - i(\theta) b(\theta)$$

Firth's Adjusted Score Function

Parametric setting: $U(\theta) = \nabla \ell(\theta) \qquad \ell(\theta) = \log \mathcal{L}(\theta)$

MLE: $\hat{\theta}$ **with bias** $b(\theta)$

Firth's idea: reduce bias of $\hat{\theta}$ **by adjusting** $U(\theta)$

$$U^{\star}(\theta) = U(\theta) - i(\theta) b(\theta)$$

Binomial logistic regression - adjust the data

Firth's Adjusted Score Function

Parametric setting: $U(\theta) = \nabla \ell(\theta) \qquad \ell(\theta) = \log \mathcal{L}(\theta)$

MLE: $\hat{\theta}$ **with bias** $b(\theta)$

Firth's idea: reduce bias of $\hat{\theta}$ **by adjusting** $U(\theta)$

$$U^{\star}(\theta) = U(\theta) - i(\theta) b(\theta)$$

Binomial logistic regression - adjust the data

Cox & Reid (1372), Barndorff-Nielsen. (561),
McCullagh and Tibshirani (261), Firth (2337)

Adjusting the Score Function

$$U^{\star}(\theta) = U(\theta) - i(\theta)b(\theta)$$

Adjusting the Score Function

$$U^\star(\theta) = U(\theta) - i(\theta)b(\theta)$$

Bias: $b(\theta) = \mathbf{E}[\hat{\theta} - \theta]$

Adjusting the Score Function

$$U^\star(\theta) = U(\theta) - i(\theta)b(\theta)$$

Bias: $b(\theta) = \mathbf{E}[\hat{\theta} - \theta]$

$$\hat{\theta} - \theta = \frac{U(\theta)}{i(\theta)} + \frac{1}{2}\left[\frac{U(\theta)}{i(\theta)}\right]^2 \frac{\ddot{U}(\theta)}{i(\theta)}$$

Adjusting the Score Function

$$U^\star(\theta) = U(\theta) - i(\theta)b(\theta)$$

Bias: $b(\theta) = \mathbf{E}[\hat{\theta} - \theta]$

$$\hat{\theta} - \theta = \frac{U(\theta)}{i(\theta)} + \frac{1}{2}\left[\frac{U(\theta)}{i(\theta)}\right]^2 \frac{\ddot{U}(\theta)}{i(\theta)}$$

Standard: both terms contribute to bias

Adjusting the Score Function

$$U^\star(\theta) = U(\theta) - i(\theta)b(\theta)$$

Bias: $b(\theta) = \mathbf{E}[\hat{\theta} - \theta]$

$$\hat{\theta} - \theta = \frac{U(\theta)}{i(\theta)} + \frac{1}{2}\left[\frac{U(\theta)}{i(\theta)}\right]^2 \frac{\ddot{U}(\theta)}{i(\theta)}$$

Standard: both terms contribute to bias

Nonparametrics: second term is of lower order

Adjusting the Score Function

$$U^\star(\theta) \quad = \quad U(\theta) - i(\theta) \cdot \mathbf{E}\left[\frac{U(\theta)}{i(\theta)}\right]$$

Adjusting the Score Function

$$U^{\star}(\theta) = U(\theta) - i(\theta) \cdot \mathbf{E}\left[\frac{U(\theta)}{i(\theta)}\right]$$

$$\approx U(\theta) - \mathbf{E}\left[U(\theta)\right]$$

Adjusting the Score Function

$$U^{\star}(\theta) = U(\theta) - i(\theta) \cdot \mathbf{E}\left[\frac{U(\theta)}{i(\theta)}\right]$$

$$\approx U(\theta) - \mathbf{E}\left[U(\theta)\right]$$

U will be a local likelihood score function

Adjusting the Score Function

$$U^{\star}(\theta) = U(\theta) - i(\theta) \cdot \mathbf{E}\left[\frac{U(\theta)}{i(\theta)}\right]$$

$$\approx U(\theta) - \mathbf{E}\left[U(\theta)\right]$$

U will be a local likelihood score function

Does this lead to an adjustment of the data?

Nadaraya-Watson Estimator

$$Y_i = g(X_i) + e_i \qquad \theta_x = g(x)$$

Nadaraya-Watson Estimator

$$Y_i = g(X_i) + e_i \qquad \theta_x = g(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{Y_i - \theta_x\}^2$$

Nadaraya-Watson Estimator

$$Y_i = g(X_i) + e_i \qquad \theta_x = g(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{Y_i - \theta_x\}^2$$

$$U(\theta_x) = -2 \sum K_h(X_i - x)\{Y_i - \theta_x\}$$

Nadaraya-Watson Estimator

$$Y_i = g(X_i) + e_i \qquad \theta_x = g(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{Y_i - \theta_x\}^2$$

$$U(\theta_x) = -2\sum K_h(X_i - x)\{Y_i - \theta_x\}$$

$$\hat{\theta}_x = \hat{g}(x) = \frac{\sum_i K_h(X_i - x)\, Y_i}{\sum K_h(X_i - x)}$$

Nadaraya-Watson Estimator

$$Y_i = g(X_i) + e_i \qquad \theta_x = g(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{Y_i - \theta_x\}^2$$

$$U(\theta_x) = -2 \sum K_h(X_i - x)\{Y_i - \theta_x\}$$

$$\hat{\theta}_x = \hat{g}(x) = \frac{\sum_i K_h(X_i - x) Y_i}{\sum K_h(X_i - x)}$$

$$\mathbf{E}\left[U(\theta_x)\right] \approx -nh^2 \sigma_K^2 \ddot{g}(x)$$

Nadaraya-Watson Estimator

$$U^\star(\theta_x) = U(\theta_x) + nh^2\sigma_K^2\ddot{g}(x)$$

Nadaraya-Watson Estimator

$$
\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} + nh^2\sigma_K^2\ddot{g}(x)
\end{aligned}
$$

Nadaraya-Watson Estimator

$$\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} \\
&\qquad + h^2\sigma_K^2\sum K_h(X_i - x)\ddot{g}(X_i)
\end{aligned}$$

Nadaraya-Watson Estimator

$$
\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} \\
&\qquad\qquad + h^2\sigma_K^2\sum K_h(X_i - x)\ddot{g}(X_i) \\
&= -2\sum K_h(X_i - x)\{Y_i - \tfrac{1}{2}h^2\sigma_K^2\ddot{g}(X_i) - \theta_x\}
\end{aligned}
$$

Nadaraya-Watson Estimator

$$
\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} + nh^2\sigma_K^2\ddot{g}(x) \\
&= -2\sum K_h(X_i - x)\{Y_i - \theta_x\} \\
&\qquad\qquad + h^2\sigma_K^2 \sum K_h(X_i - x)\ddot{g}(X_i) \\
&= -2\sum K_h(X_i - x)\{Y_i - \tfrac{1}{2}h^2\sigma_K^2\ddot{g}(X_i) - \theta_x\} \\
&= -2\sum K_h(X_i - x)\{Y_i^\star - \theta_x\}
\end{aligned}
$$

Ideal Data Sharpening

$$Y_i^\star = Y_i - \tfrac{1}{2} h^2 \sigma_K^2 \ddot{g}(X_i) \qquad \textbf{Ideal Data Sharpening}$$

Ideal Data Sharpening

$$Y_i^\star = Y_i - \tfrac{1}{2}h^2\sigma_K^2\ddot{g}(X_i) \qquad \textbf{Ideal Data Sharpening}$$

$$\widehat{Y}_i^\star = Y_i - [\hat{g}(X_i) - Y_i] \qquad \textbf{Data Sharpening (Hall)}$$

Kernel Density Estimation

$$X_i \sim f \qquad \theta_x = f(x)$$

Kernel Density Estimation

$$X_i \sim f \qquad \theta_x = f(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\theta_x - n \int K(u - x) e^{\theta_x} \, du$$

Kernel Density Estimation

$$X_i \sim f \qquad \theta_x = f(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\theta_x - n \int K(u - x) e^{\theta_x} du$$

$$U(\theta_x) = \sum \{K_h(X_i - x) - e^{\theta_x}\}$$

Kernel Density Estimation

$$X_i \sim f \qquad \theta_x = f(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\theta_x - n\int K(u-x)\,e^{\theta_x}\,du$$

$$U(\theta_x) = \sum\{K_h(X_i - x) - e^{\theta_x}\}$$

$$\hat{\theta}_x = \log\left[\hat{f}(x)\right] \qquad \hat{f}(x) = \tfrac{1}{n}\sum K_h(X_i - x)$$

Kernel Density Estimation

$$X_i \sim f \qquad \theta_x = f(x)$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\theta_x - n \int K(u-x)\, e^{\theta_x}\, du$$

$$U(\theta_x) = \sum \{ K_h(X_i - x) - e^{\theta_x} \}$$

$$\hat{\theta}_x = \log \left[\hat{f}(x) \right] \qquad \hat{f}(x) = \tfrac{1}{n} \sum K_h(X_i - x)$$

$$\mathbf{E}[\, U(\theta_x) \,] \approx \tfrac{1}{2} n h^2 \sigma_K^2 \ddot{f}(x)$$

Kernel Density Estimation

$$U^\star(\theta_x) = U(\theta_x) - \frac{1}{2} n h^2 \sigma_K^2 \ddot{f}(x)$$

Kernel Density Estimation

$$
\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) - \tfrac{1}{2} n h^2 \sigma_K^2 \ddot{f}(x) \\
&= \sum \left\{ \left[K_h(X_i - x) - h_s^2 \sigma_K^2 \ddot{f}(x) \right] - e^{\theta_x} \right\}
\end{aligned}
$$

Kernel Density Estimation

$$
\begin{aligned}
U^{\star}(\theta_x) &= U(\theta_x) - \tfrac{1}{2} n h^2 \sigma_K^2 \ddot{f}(x) \\
&= \sum \left\{ \left[K_h(X_i - x) - h_s^2 \sigma_K^2 \ddot{f}(x) \right] - e^{\theta_x} \right\} \\
&\approx \sum \left\{ K_h \left(\left[X_i + h_s^2 \sigma_K^2 \tfrac{\dot{f}(X_i)}{f(X_i)} \right] - x \right) - e^{\theta_x} \right\}
\end{aligned}
$$

Kernel Density Estimation

$$
\begin{aligned}
U^\star(\theta_x) &= U(\theta_x) - \tfrac{1}{2} n h^2 \sigma_K^2 \ddot{f}(x) \\
&= \sum \left\{ \left[K_h(X_i - x) - h_s^2 \sigma_K^2 \ddot{f}(x) \right] - e^{\theta_x} \right\} \\
&\approx \sum \left\{ K_h \left(\left[X_i + h_s^2 \sigma_K^2 \frac{\dot{f}(X_i)}{f(X_i)} \right] - x \right) - e^{\theta_x} \right\} \\
&= \sum \{ K_h(X_i^\star - x) - e^{\theta_x} \}
\end{aligned}
$$

Kernel Density Estimation

$$X_i^\star = X_i + h_s^2 \sigma_K^2 \frac{\dot{f}(X_i)}{f(X_i)}$$

Ideal Data Sharpening

Kernel Density Estimation

$$X_i^\star = X_i + h_s^2 \sigma_K^2 \frac{\dot{f}(X_i)}{f(X_i)}$$

Ideal Data Sharpening

$$\hat{X}_i^\star = X_i + \frac{\sum_j K_{h_s}(X_j - X_i)(X_j - X_i)}{\sum_j K_{h_s}(X_j - X_i)}$$

Kernel Density Estimation

$$X_i^\star = X_i + h_s^2 \sigma_K^2 \frac{\dot{f}(X_i)}{f(X_i)}$$ **Ideal Data Sharpening**

$$\hat{X}_i^\star = X_i + \frac{\sum_j K_{h_s}(X_j - X_i)(X_j - X_i)}{\sum_j K_{h_s}(X_j - X_i)}$$

$$= \frac{\sum_j K_{h_s}(X_j - X_i) X_j}{\sum_j K_{h_s}(X_j - X_i)}$$ **Data Sharpening (Hall)**

$$X_i^\star = X_i + h_s^2 \sigma_K^2 \frac{\dot{f}(X_i)}{f(X_i)}$$ **Ideal Data Sharpening**

$$\hat{X}_i^\star = X_i + \frac{\sum_j K_{h_s}(X_j - X_i)(X_j - X_i)}{\sum_j K_{h_s}(X_j - X_i)}$$

$$= \frac{\sum_j K_{h_s}(X_j - X_i) X_j}{\sum_j K_{h_s}(X_j - X_i)}$$ **Data Sharpening (Hall)**

Proof of concept

Derivative Estimation

$$Y_i = g(X_i) + e_i \qquad \theta_{1x} = g(x),\ \theta_{2x} = \dot{g}(x)$$

Derivative Estimation

$$Y_i = g(X_i) + e_i \qquad \theta_{1x} = g(x),\ \theta_{2x} = \dot{g}(x)$$

$$\ell(g, x) = \sum K_h(X_i - x)\{Y_i - g(X_i)\}^2$$

Derivative Estimation

$$Y_i = g(X_i) + e_i \qquad \theta_{1x} = g(x),\ \theta_{2x} = \dot{g}(x)$$

$$\ell(g, x) = \sum K_h(X_i - x)\{\,Y_i - g(X_i)\}^2$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{\,Y_i - \theta_{1x} - \theta_{2x}(X_i - x)\}^2$$

Derivative Estimation

$$Y_i = g(X_i) + e_i \qquad \theta_{1x} = g(x),\ \theta_{2x} = \dot{g}(x)$$

$$\ell(g, x) = \sum K_h(X_i - x)\{Y_i - g(X_i)\}^2$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{Y_i - \theta_{1x} - \theta_{2x}(X_i - x)\}^2$$

$$U(\theta_{2x}) = -2\sum K_h(X_i - x)(X_i - x)\{Y_i - \theta_{1x} - \theta_{2x}(X_i - x)\}$$

Derivative Estimation

$$Y_i = g(X_i) + e_i \qquad \theta_{1x} = g(x),\ \ \theta_{2x} = \dot{g}(x)$$

$$\ell(g,x) = \sum K_h(X_i - x)\{\,Y_i - g(X_i)\}^2$$

$$\ell(\theta_x) = \sum K_h(X_i - x)\{\,Y_i - \theta_{1x} - \theta_{2x}(X_i - x)\}^2$$

$$U(\theta_{2x}) = -2\sum K_h(X_i - x)(X_i - x)\{\,Y_i - \theta_{1x} - \theta_{2x}(X_i - x)\}$$

$$\mathbf{E}\left[U(\theta_{2x})\right] \approx -\tfrac{nh^4\mu_4^K}{3}\,\dddot{g}(x)$$

Derivative Estimation

$$U^\star(\theta_{2x}) = U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x)$$

Derivative Estimation

$$U^\star(\theta_{2x}) = U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x)$$

$$= -2\sum K_h(X_i - x)(X_i - x)\,Y_i + \frac{nh^4\mu_4^K}{3}\dddot{g}(x)$$

Derivative Estimation

$$
\begin{aligned}
U^\star(\theta_{2x}) &= U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i
\end{aligned}
$$

Derivative Estimation

$$
\begin{aligned}
U^\star(\theta_{2x}) &= U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i \\
&\qquad + \frac{h^2\mu_4^K}{3\sigma_K^2}\sum K_h(X_i - x)(X_i - x)\ddot{g}(X_i)
\end{aligned}
$$

Derivative Estimation

$$
\begin{aligned}
U^\star(\theta_{2x}) &= U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i \\
&\qquad + \frac{h^2\mu_4^K}{3\sigma_K^2}\sum K_h(X_i - x)(X_i - x)\ddot{g}(X_i) \\
&= -2\sum K_h(X_i - x)(X_i - x)\left\{ Y_i - \frac{h^2\mu_4^K}{6\sigma_K^2}\ddot{g}(X_i)\right\}
\end{aligned}
$$

Derivative Estimation

$$
\begin{aligned}
U^{\star}(\theta_{2x}) &= U(\theta_{2x}) + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i + \frac{nh^4\mu_4^K}{3}\dddot{g}(x) \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i \\
&\qquad + \frac{h^2\mu_4^K}{3\sigma_K^2}\sum K_h(X_i - x)(X_i - x)\ddot{g}(X_i) \\
&= -2\sum K_h(X_i - x)(X_i - x)\left\{Y_i - \frac{h^2\mu_4^K}{6\sigma_K^2}\ddot{g}(X_i)\right\} \\
&= -2\sum K_h(X_i - x)(X_i - x)\,Y_i^{\star}
\end{aligned}
$$

Derivative Estimation

$$Y_i^\star = Y_i - \frac{h^2 \mu_4^K}{6\sigma_K^2} \ddot{g}(X_i)$$

Ideal Data Sharpening

Derivative Estimation

$$Y_i^\star = Y_i - \frac{h^2\mu_4^K}{6\sigma_K^2}\ddot{g}(X_i)$$ **Ideal Data Sharpening**

$$\widehat{Y}_i^\star = Y_i - \frac{\mu_4^K}{3\sigma_K^4}\left[\hat{g}(X_i) - Y_i\right]$$ **Data Sharpening (BBS)**

$$Y_i^\star = Y_i - \frac{h^2\mu_4^K}{6\sigma_K^2}\ddot{g}(X_i)$$ **Ideal Data Sharpening**

$$\widehat{Y}_i^\star = Y_i - \frac{\mu_4^K}{3\sigma_K^4}\left[\hat{g}(X_i) - Y_i\right]$$ **Data Sharpening (BBS)**

Nadaraya-Watson Estimator

Derivative Estimation

$$Y_i^\star = Y_i - \frac{h^2 \mu_4^K}{6\sigma_K^2} \ddot{g}(X_i)$$ **Ideal Data Sharpening**

$$\widehat{Y}_i^\star = Y_i - \frac{\mu_4^K}{3\sigma_K^4} \left[\hat{g}(X_i) - Y_i\right]$$ **Data Sharpening (BBS)**

Nadaraya-Watson Estimator

$$Y_i^\star = Y_i - \frac{1}{2} h^2 \sigma_K^2 \ddot{g}(X_i)$$ **Ideal Data Sharpening**

Derivative Estimation

$$Y_i^\star = Y_i - \frac{h^2 \mu_4^K}{6\sigma_K^2} \ddot{g}(X_i)$$ **Ideal Data Sharpening**

$$\widehat{Y}_i^\star = Y_i - \frac{\mu_4^K}{3\sigma_K^4} \left[\hat{g}(X_i) - Y_i\right]$$ **Data Sharpening (BBS)**

Nadaraya-Watson Estimator

$$Y_i^\star = Y_i - \frac{1}{2} h^2 \sigma_K^2 \ddot{g}(X_i)$$ **Ideal Data Sharpening**

$$\widehat{Y}_i^\star = Y_i - \left[\hat{g}(X_i) - Y_i\right]$$ **Data Sharpening (Hall)**

Gas mileage Y as a function g of cruising speed X

Fuel Efficiency

Gas mileage Y as a function g of cruising speed X

Where does the first derivative $\dot{g}$ equal zero?

Fuel Efficiency

Gas mileage Y **as a function** g **of cruising speed** X

Where does the first derivative $\dot{g}$ **equal zero?**

Usual local linear derivative estimate versus

Fuel Efficiency

Gas mileage Y as a function g of cruising speed X

Where does the first derivative $\dot{g}$ equal zero?

Usual local linear derivative estimate versus

Sharpened counterpart

Fuel Efficiency

Gas mileage Y as a function g of cruising speed X

Where does the first derivative $\dot{g}$ equal zero?

Usual local linear derivative estimate versus

Sharpened counterpart

Data:

Gas mileage Y as a function g of cruising speed X

Where does the first derivative $\dot{g}$ equal zero?

Usual local linear derivative estimate versus

Sharpened counterpart

Data:

15 observations of gas mileage

Fuel Efficiency

Gas mileage Y as a function g of cruising speed X

Where does the first derivative $\dot{g}$ equal zero?

Usual local linear derivative estimate versus

Sharpened counterpart

Data:

15 observations of gas mileage
Equally spaced cruising speeds [5,75]

Fuel Efficiency

Small simulation

Model:

$$g(x) = 25\sin(x/30) + 5 + e$$

Small simulation

Model:

$$g(x) = 25\sin(x/30) + 5 + e$$

Optimal speed is 47.12 & 1000 samples

Small simulation

Model:

$$g(x) = 25\sin(x/30) + 5 + e$$

Optimal speed is 47.12 & 1000 samples

Standard: Bias of 9.79 & root MSE of 11.41

Small simulation

Model:

$$g(x) = 25\sin(x/30) + 5 + e$$

Optimal speed is 47.12 & 1000 samples

Standard: Bias of 9.79 & root MSE of 11.41

Sharpened: Bias of 6.41 & root MSE of 7.92

Derivative estimation

Derivative estimation

Firth's adjusted score function ideal vehicle

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data
Actual DS unclear - residuals over correct for bias

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data
Actual DS unclear - residuals over correct for bias

Derivatives in a multivariate setting (spatial)

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data
Actual DS unclear - residuals over correct for bias

Derivatives in a multivariate setting (spatial)

A single adjustment may/will not bias correct in all directions

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits

"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data

Actual DS unclear - residuals over correct for bias

Derivatives in a multivariate setting (spatial)

A single adjustment may/will not bias correct in all directions

Consider a very regular grid

Expanding the boundaries of DS

Derivative estimation

Firth's adjusted score function ideal vehicle

Higher order derivatives

Complex coefficients & residuals from higher order polynomial fits
"Simple" structure but laborious calculations - computer algebra

GLM

Clear how to "ideally" sharpen the data
Actual DS unclear - residuals over correct for bias

Derivatives in a multivariate setting (spatial)

A single adjustment may/will not bias correct in all directions
Consider a very regular grid
Consider a single direction