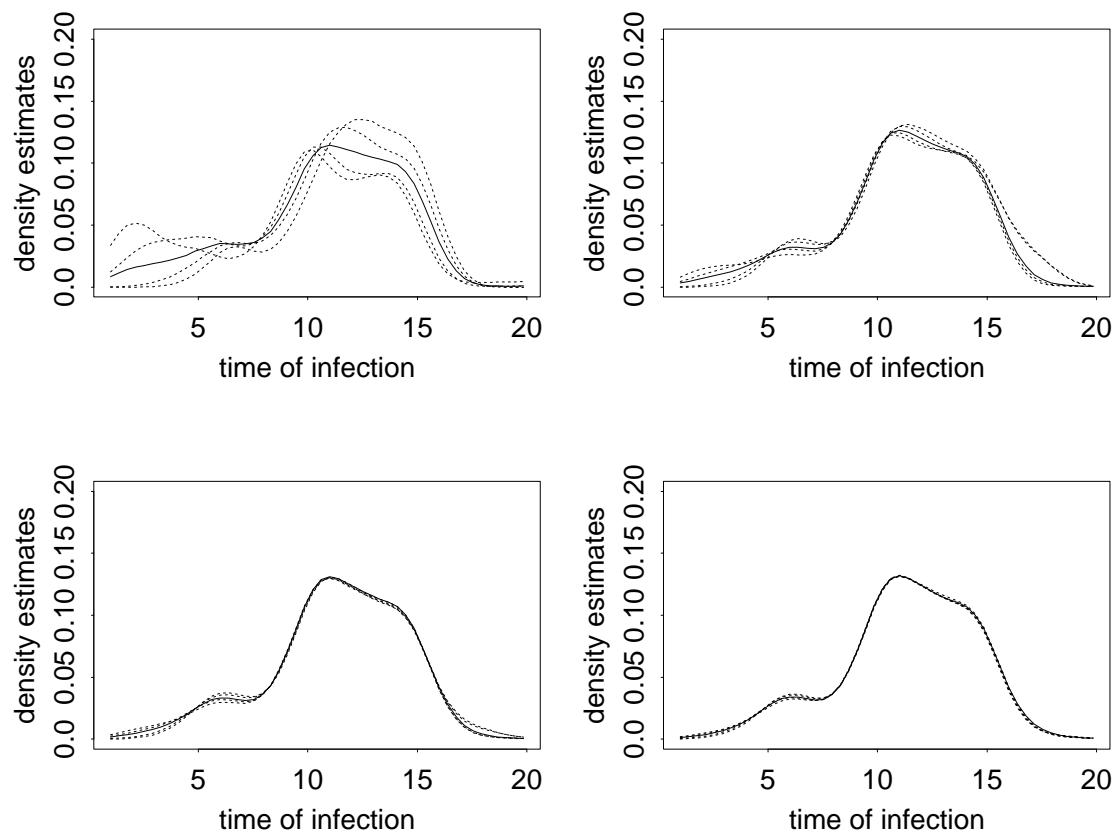


ICE: Iterated Conditional Expectations



Summary

Local Likelihood Density estimation for Interval Censored Data

- A Kernel Density Estimate
- Relationship to Self-Consistency
- Relationship to a Local Likelihood Class and the EM
- Implementation: A Newton Iteration
- For the Class: Explicit Solution for the M-step
- Symbolic Newton Raphson

Iterated Conditional Expectations (ICE)

- Other Local Likelihood Classes
- More Symbolic Computation

The Data Type

- Observe $(L_i, R_i) = I_i$
- X_i Unobserved $X_i \in I_i$
- Individuals at risk are periodically tested for HIV
- L_i time of last negative test
- R_i time of first positive test
- X_i Time of infection
- Histogram: (L_i, R_i) Bin Containing X_i
- Often $X_i \sim Unif(L_i, R_i)$

A Kernel Density Estimate

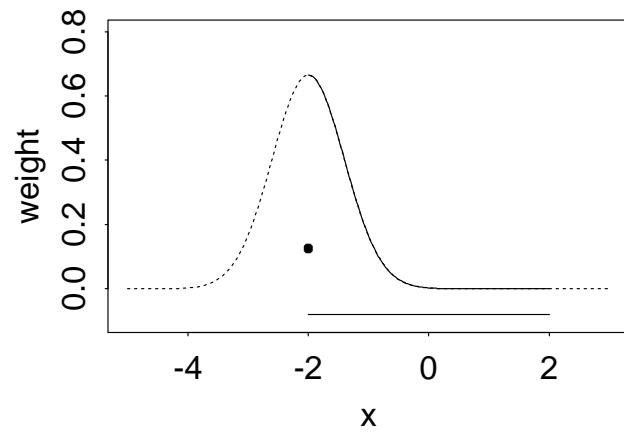
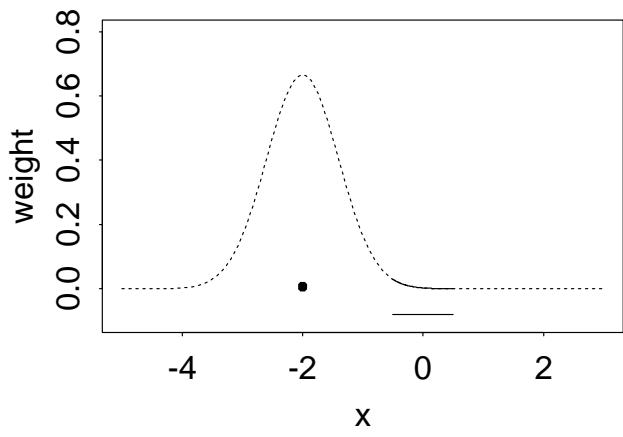
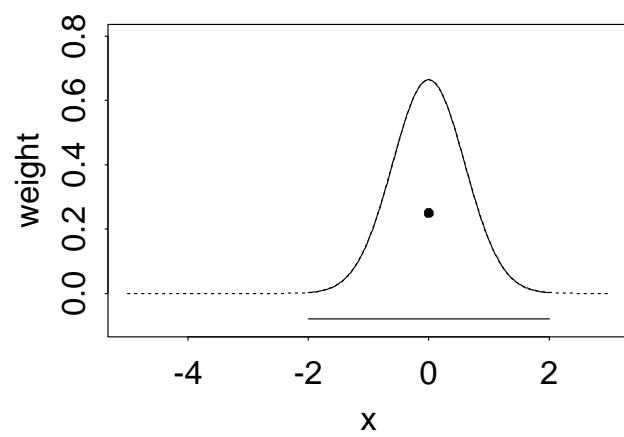
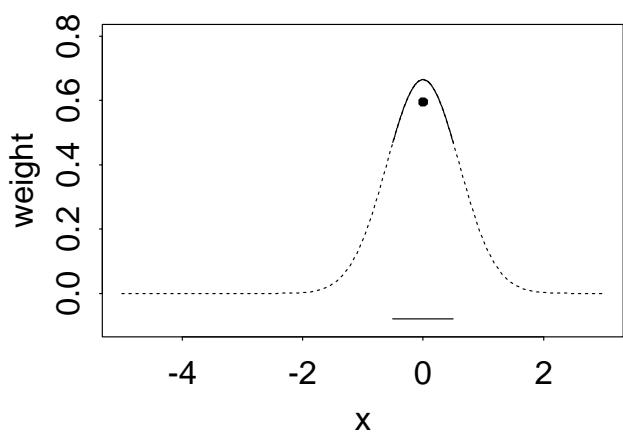
$$\hat{f}_c(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

$$w_i = \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

$$w_i = E\left[\frac{1}{h} K\left(\frac{X_i - x}{h}\right) \mid I_i\right]$$

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n E\left[\frac{1}{h} K\left(\frac{X_i - x}{h}\right) \mid I_i\right]$$

Goutis (1997)



An iteration

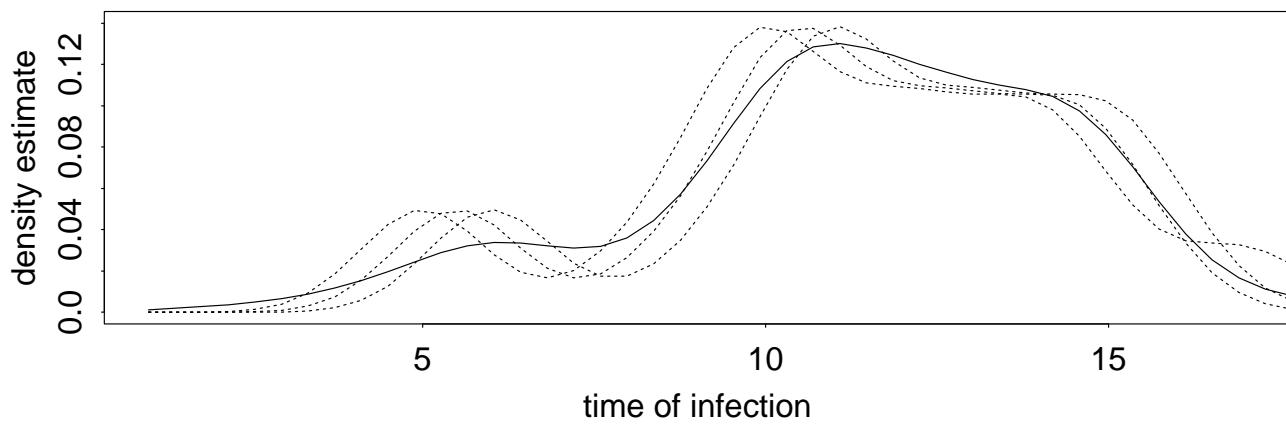
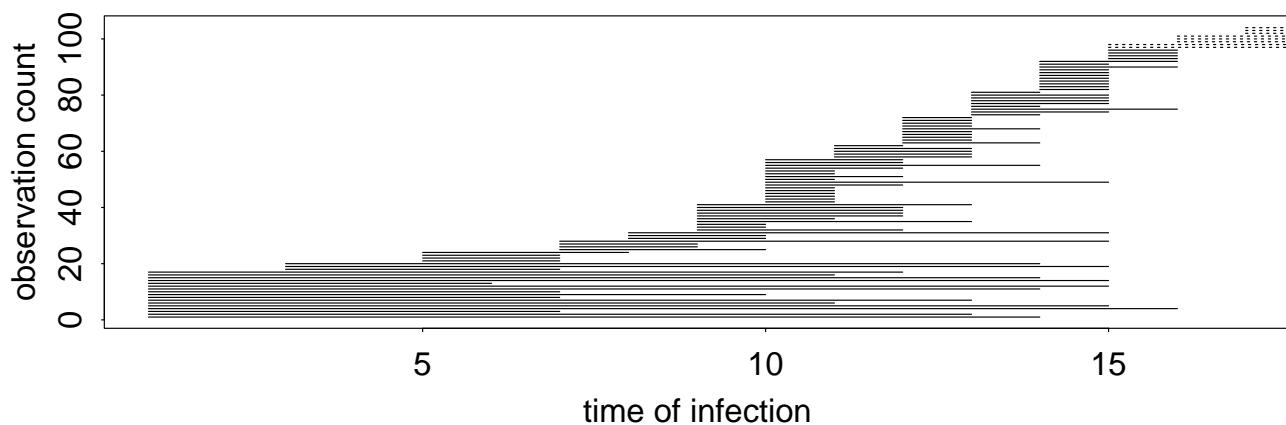
$$E[\cdot | I_i] = ?$$

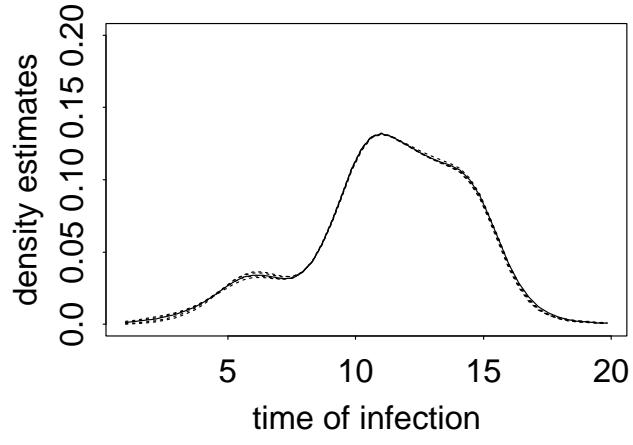
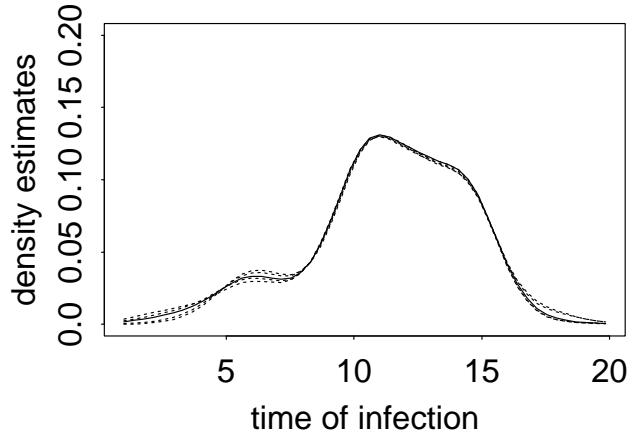
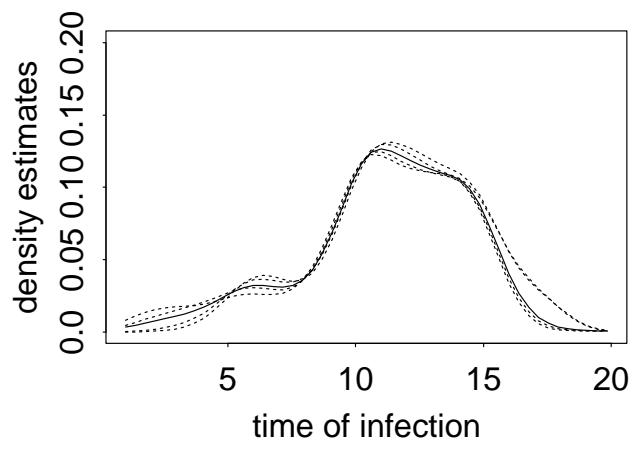
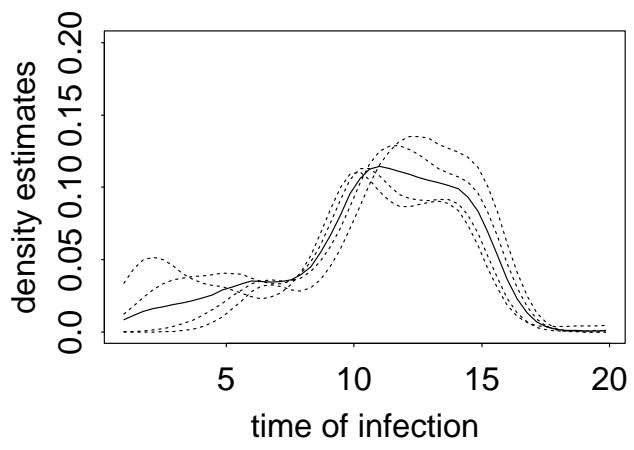
$$\hat{f}_j(x) = \frac{1}{n} \sum_{i=1}^n E_{\hat{f}_{j-1}} \left[\frac{1}{h} K \left(\frac{X_i - x}{h} \right) \middle| I_i \right]$$

\hat{f}_{j-1} is normalized over each I_i

A fixed point equation

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n E_{\hat{f}} \left[\frac{1}{h} K \left(\frac{X_i - x}{h} \right) \middle| I_i \right]$$





- Initial value: a member of the Beta location scale family

$$\hat{F}_c(x)=\int_{-\infty}^x \hat{f}_c(u)du,$$

$$\lim_{h\downarrow 0}\hat{F}_c(x)=F_n$$

$$\hat{F}_j(x)=\int_{-\infty}^x \hat{f}_j(u)du,$$

$$\lim_{h\downarrow 0}\hat{F}_j(x)=?$$

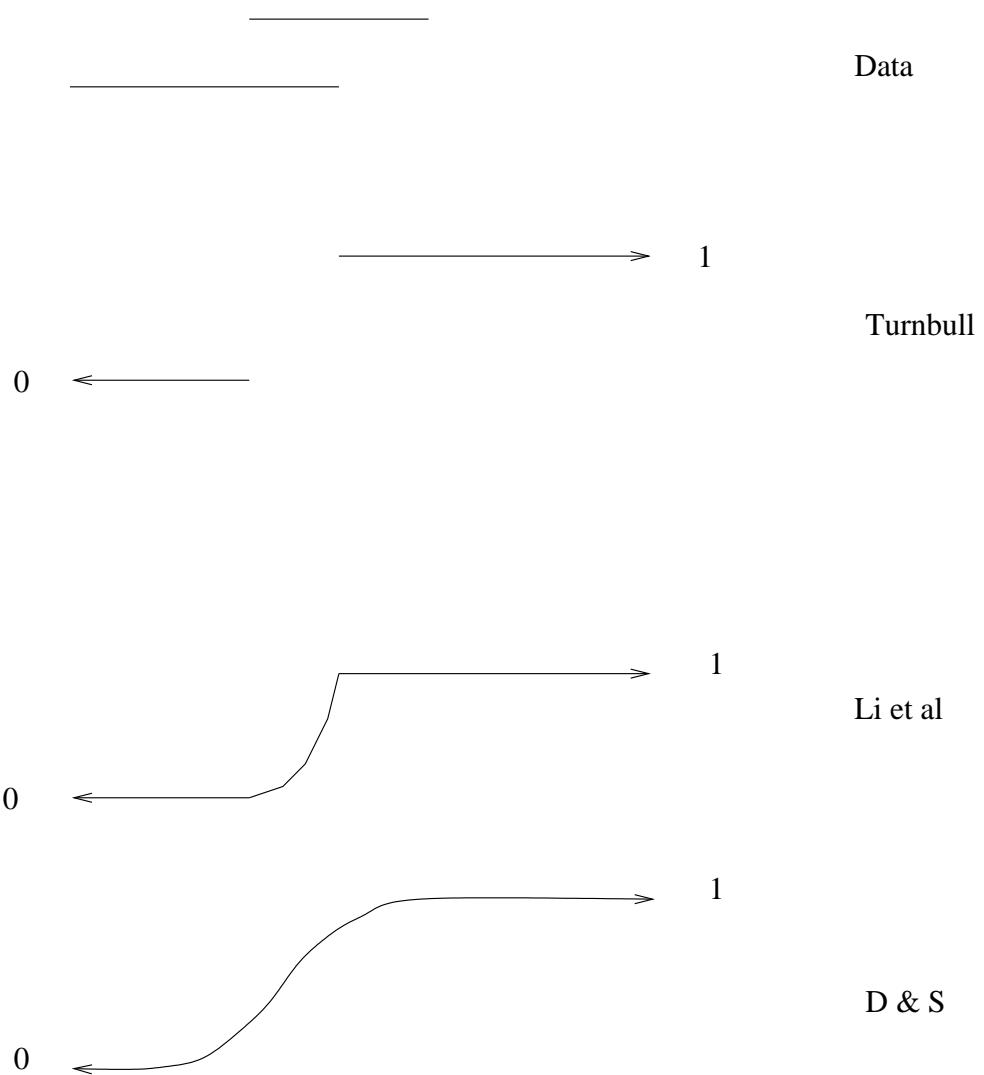
$$9\\$$

Self-Consistency

- F_n is the NPMLE of F for complete data
- F_k is the NPMLE of F for right censored data
- F_t is the NPMLE of F for interval censored data
- $F_n \equiv (X_i, 1/n), \quad F_k \equiv (F_i, w_i) \quad F_t \equiv (J_r, p_r)$
- $J_r \leftarrow$ innermost intervals
- Efron (1967), Turnbull (1976), Li, Watkins and Yu (1997)
- Self-Consistency

$$\check{F}_j(x) = E_{j-1} [F_n(x) | I_i \forall i].$$

Innermost Intervals



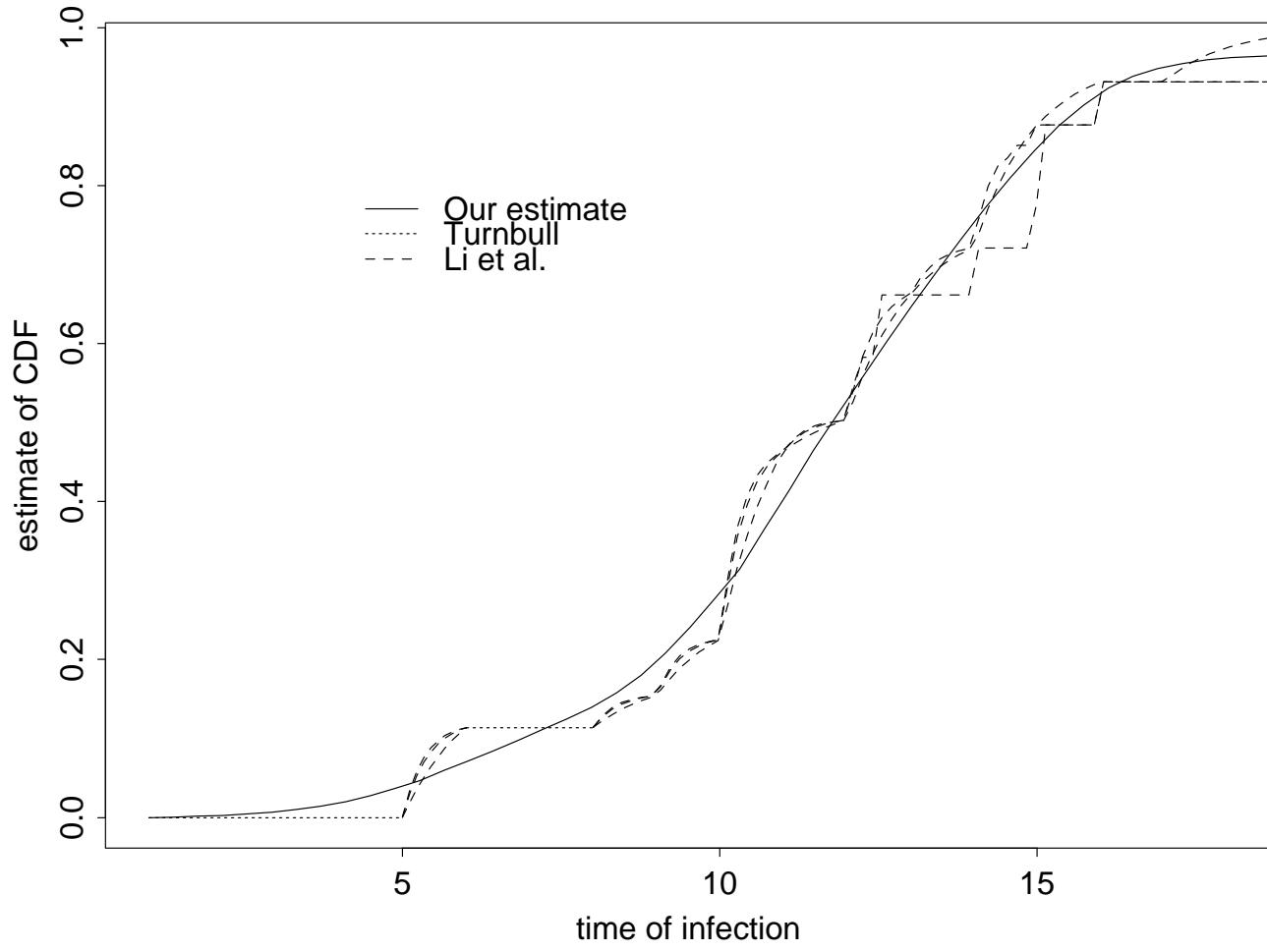
Relationship to Self-Consistency

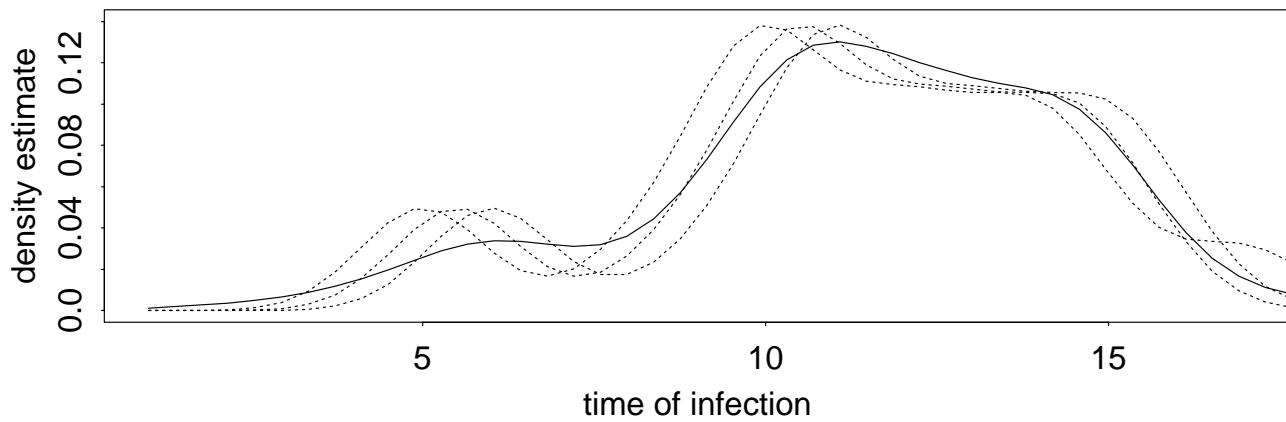
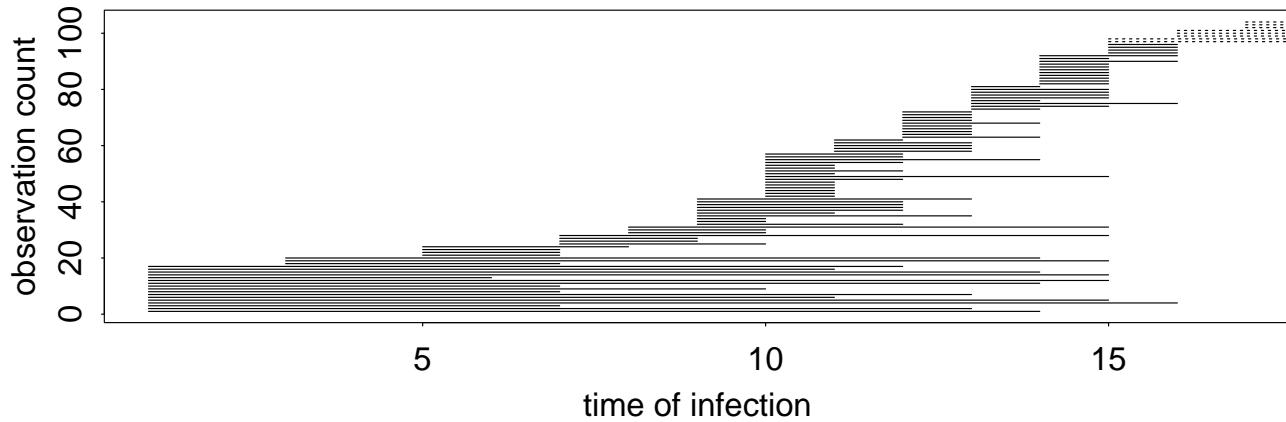
Theorem: $\lim_{h \downarrow 0} \hat{F}_j(x) = \check{F}_j(x)$

Proof:

$$\check{F}_j(x) = \frac{1}{n} \sum_{i=1}^n E_{j-1} [I[X_i \leq x] | I_i]$$

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n E_{j-1} \left[\int_{-\infty}^x \frac{1}{h} K\left(\frac{X_i - u}{h}\right) du \middle| I_i \right]$$





$$\hat{f}_t(x) = \sum_{r=1}^R p_r \frac{1}{h} K\left(\frac{x - ?}{h}\right)$$

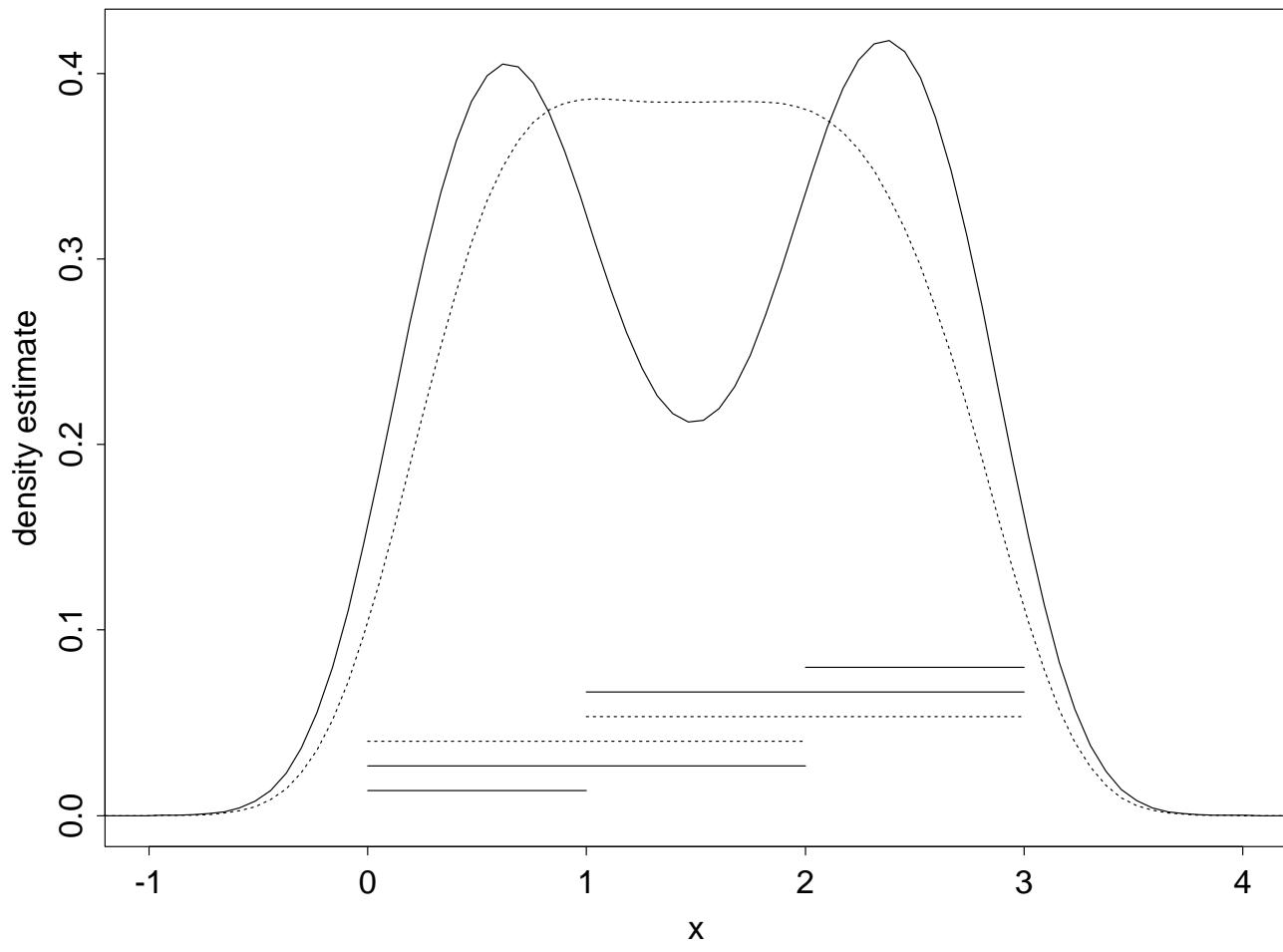
$$(0, 1), (0, 2), (0, 2), (1, 3), (1, 3), (2, 3)$$

$$\hat{F}_t$$

$$(0, 1), (1, 2), (2, 3) \text{ with weights } \frac{1}{3}, \frac{1}{3}, \frac{1}{3}$$

$$\frac{1}{2}, 0, \frac{1}{2}$$

$$\text{reduced data } (0, 1), (0, 2), (1, 3), (2, 3)$$



Implementation

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\int_{I_i} K_h(x-y) \hat{f}(y) dy}{\int_{I_i} \hat{f}(y) dy} \right\}$$

- $\mathcal{M} = \{x_k\}_{k=1}^M \quad f = (f^1, \dots, f^M)^T$ with $f^k = f(x_k)$
- Trapezoidal quadrature rule

$$f_k = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\sum_{l: x_l \in I_i} K_h(x_l - x_k) f_l}{\sum_{l: x_l \in I_i} f_l} \right\}.$$

$$\hat{f} = \mathcal{G}(\tilde{f})$$

$$\hat{f} = \mathcal{G}(\hat{f}) \iff \hat{f}_{j+1} = \mathcal{G}(\hat{f}_j)$$

- Convergence to a unique solution for sufficiently large “h”

Newton Iteration

- Alternatively consider solving $\mathcal{U}(\hat{\tilde{f}}) = 0$

$$\mathcal{U}(\tilde{f}) = \tilde{f} - \mathcal{G}(\tilde{f})$$

- Construct a Newton iteration as

$$\begin{aligned}\tilde{f}_{j+1} &= \tilde{\mathcal{G}}(\tilde{f}_j) = \tilde{f}_j - \left[\nabla \mathcal{U}(\tilde{f}_j) \right]^{-1} \mathcal{U}(\tilde{f}_j) \\ &\implies \hat{\tilde{f}} = \tilde{\mathcal{G}}(\hat{\tilde{f}})\end{aligned}$$

- Local convergence guaranteed provided good initial guess
- General Strategy: Iterate $\tilde{f}_{j+1} = \mathcal{G}(\tilde{f}_j)$ until $\rho(\nabla \mathcal{G}) < 1$

Empirical results

h	ρ_1	ρ_2	ρ_∞	original	Newton
.01	7.74	1.26	1.0	125	fails*
.3	1.475	.795	.800	53	fails, 4 ¹
.4	1.180	.699	.716	41	50, 3 ¹
.5	.981	.614	.633	30	3
.6	.878	.537	.588	26	3
.7	.777	.481	.551	24	3
1	.564	.364	.433	19	3
1.2	.488	.325	.368	17	3

Local Likelihood Density Estimation

- Loader (1996), Hjort and Jones (1996)
- Embed KDE in a larger local likelihood class

$$\mathcal{L}(f, x) = \sum_{i=1}^n K_h(X_i - x) \log\{f(X_i)\} - n \int_{\mathfrak{R}} K_h(u - x) f(u) du$$

- $\log\{f(u)\} = \mathcal{P}(u - x)$
- $\mathcal{P}(u - x) = a_0 + a_1(u - x) + \dots + a_p(u - x)^p$
- Solve: $\frac{\partial}{\partial \tilde{a}} \mathcal{L}(f, x) = 0 \quad \forall x$
- $\hat{f}(x) = \exp\{\hat{a}_0\}$
- Example:

$$\mathcal{P}(u - x) = a_0 \implies \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_h(X_i - x)$$

Local Likelihood Density Estimation for Interval Censored Data

$$\mathcal{L}(f, x) = \sum E[K_h(X_i - x) \log\{f(X_i)\} | I_i] - n \int_{\mathfrak{R}} K_h(u - x) f(u) du$$

- EM-type algorithm
- M-step: Solve $\frac{\partial}{\partial \hat{a}} \mathcal{L}(f, x) = 0$ to get $\hat{f}(x) = \exp\{\hat{a}_0\}$
- E-step: Compute expectations wrt $\hat{f}(x) = \exp\{\hat{a}_0\}$
- $\mathcal{P}(u - x) = a_0 \implies \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n E_{\hat{f}}[K_h(X_i - x) | I_i]$

More fixed point equations

- Gaussian kernel

- $\mathcal{P}(u - x) = a_0 + a_1(u - x) \implies$

$$\hat{f}_j^l(x) = \hat{f}_j(x) \exp \left[-\frac{h^2}{2} \left\{ \dot{\hat{f}}_j(x)/\hat{f}_j(x) \right\}^2 \right]$$

- $\mathcal{P}(u - x) = a_0 + a_1(u - x) + a_2(u - x)^2 \implies$

$$\hat{f}_j^q(x) = \hat{f}_j(x) \hat{R} \exp \left[-\frac{h^2 \hat{R}^2}{2} \left\{ \dot{\hat{f}}_j(x)/\hat{f}_j(x) \right\}^2 \right]$$

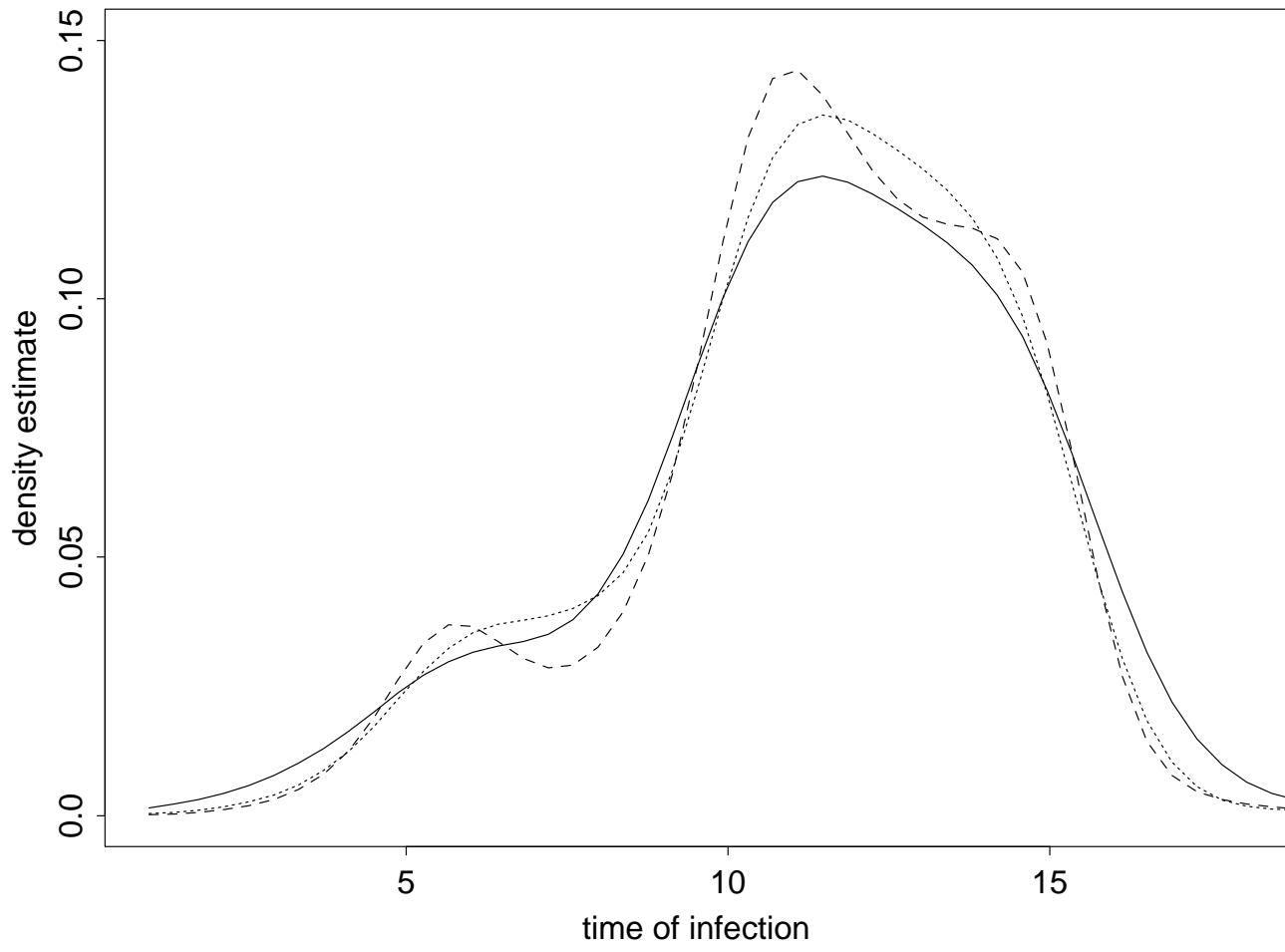
$$\dot{\hat{f}}_j(x) = \frac{\partial}{\partial x} \hat{f}_j(x) = n^{-1} \sum E_{\hat{f}_{j-1,r}} \left[\frac{\partial}{\partial x} \{K_h(X_i - x)\} | I_i \right]$$

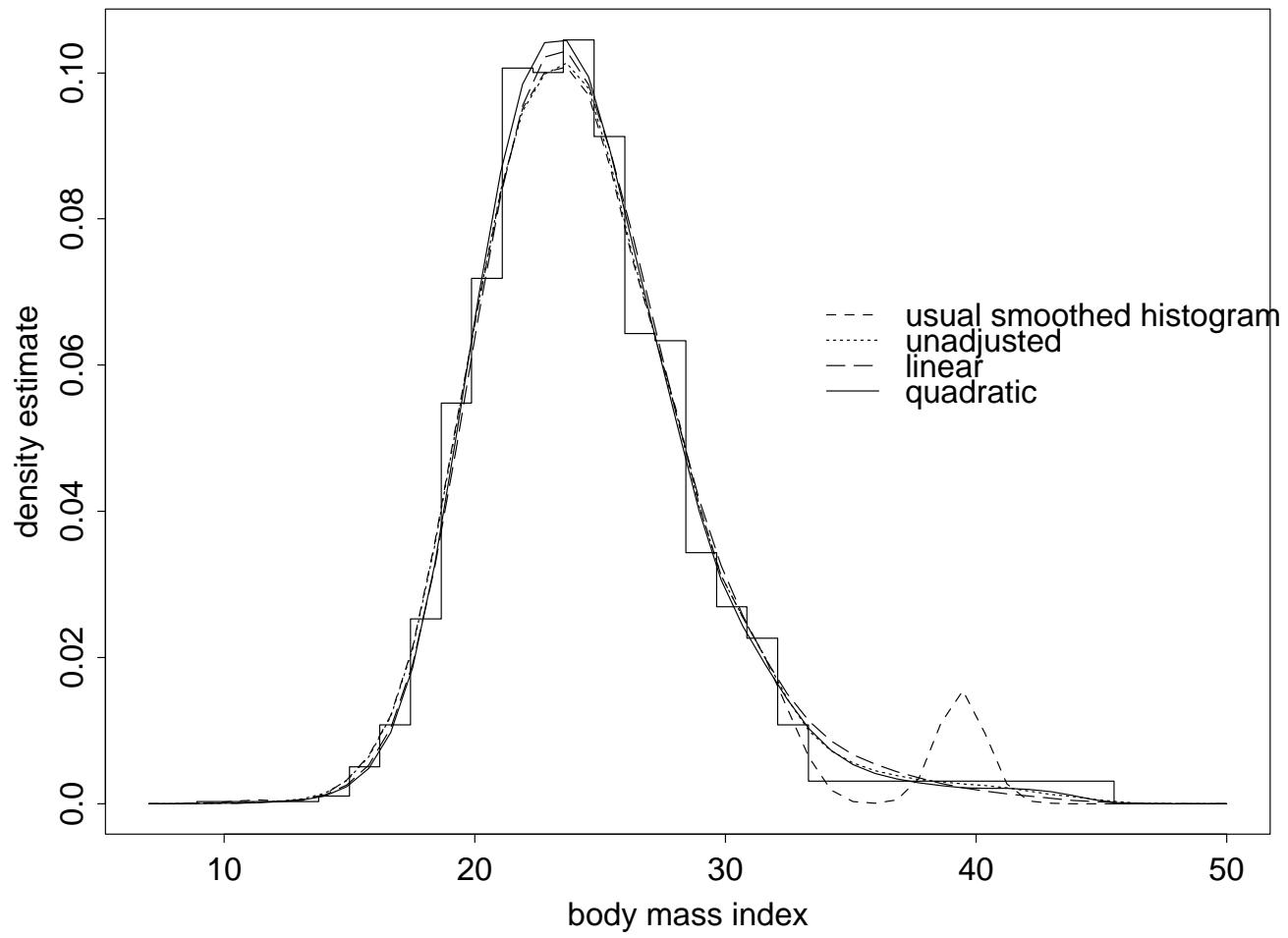
- Analogous to Hjort and Jones (1996)
- Newton Implementation for the class?

Symbolic Newton Raphson

- Newton Raphson: $\Delta_i = \Delta_{i-1}^2 \times O_p(1)$
- Symbolic Newton Raphson: $\Delta_i = \Delta_1 \Delta_{i-1} \times O_p(1)$
(Andrews and Stafford 2000)
- Difference: denominator is replaced by its leading term
- $\hat{\mathbf{a}}$ solves local likelihood equations
- $\tilde{\mathbf{a}}$ expression based on first k steps
- $\tilde{\mathbf{a}} - \hat{\mathbf{a}}$ is order $k + 1$

$$\hat{f}_j^l(x) = \hat{f}_j(x) \exp \left\{ -\frac{1}{2} \left[\left\{ \frac{\hat{f}_j}{\hat{f}_j(x)} \right\}^2 h^2 \sigma^2 + \left\{ \frac{\hat{f}_j}{4\hat{f}_j(x)} \right\}^4 h^4 (3\sigma^2 - \kappa) \right] \right\}$$





Local Log-Linear

h	ρ_1	ρ_2	ρ_∞	original	Newton
.3	1.502	.781	.858	53	fails*, 13 ¹
.4	1.242	.700	.796	51	fails*, 8
.5	1.044	.629	.756	47	4
.6	.926	.560	.730	47	fails*, 4
.7	.809	.506	.713	43	4
.8	.701	.443	.692	38	4

ICE: Other local likelihood classes

- Multivariate density estimation
- Hazard or intensity estimation (Betensky and co-authors)
- Regression
- Usefulness of Newton iteration
- Usefulness of symbolic computation
- Observations

Locally constant regression (for a GLM)

$$\mathcal{L}_x(\mathbf{a}) = \sum_{i=1}^n w_i(x) l\{Y_i, \mathcal{P}(X_i - x)\}$$

- EM-type \implies

$$\hat{g}(x) = \sum_{i=1}^n w_i(x) E_{\hat{g}}[Y_i | I_i] / \sum_{i=1}^n w_i(x)$$

- Newton Implementation \implies

$$\nabla \mathcal{G} = \sum_{i=1}^n w_i(x) \frac{V_{\hat{g}}[Y_i | I_i]}{V_{\hat{g}}[Y_i]} / \sum_{i=1}^n w_i(x)$$

Locally linear - more Symbolic Newton Raphson

- Multivariate Density Estimation:

$$\tilde{f}(x) = \hat{f}^0 \exp \left\{ -\frac{h^2 (\sum_r \hat{f}^r)^2 \mu_2}{2 \hat{f}^{0^2}} + \frac{h^4 (\sum_r \hat{f}^r)^4}{8 \hat{f}^{0^4}} (\mu_4 - 3\mu_2^2) \right\}$$

- Gaussian Regression:

$$\bar{Y}_w + (x - \bar{X}_w) \frac{\sigma_w(X, Y)}{\sigma_w(X, X)} \quad \text{vs} \quad \bar{Y}_w + (x - \bar{X}_w) \frac{\sigma_w(X, Y)}{h^2 \mu_2}$$

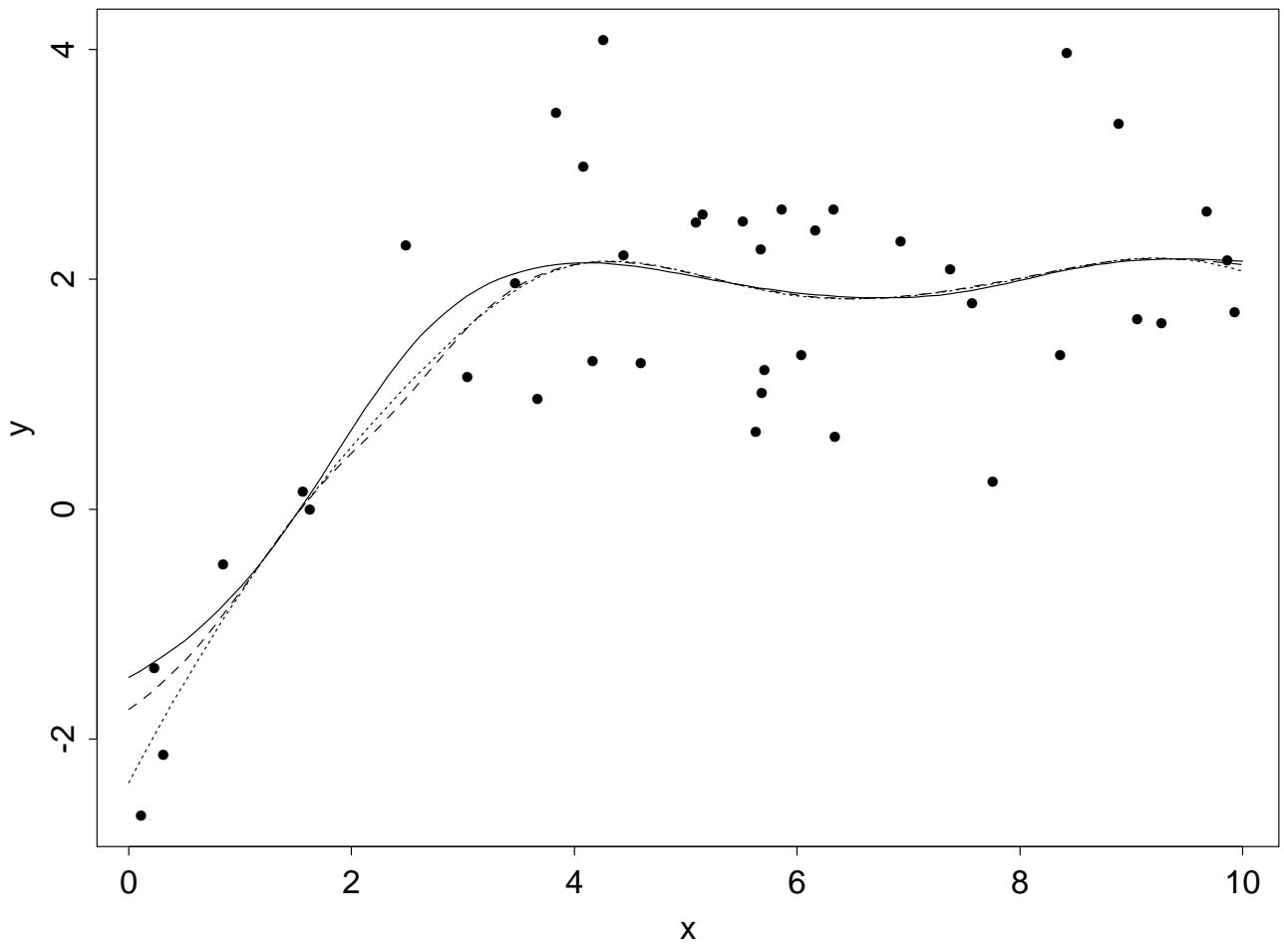
- Poisson Regression:

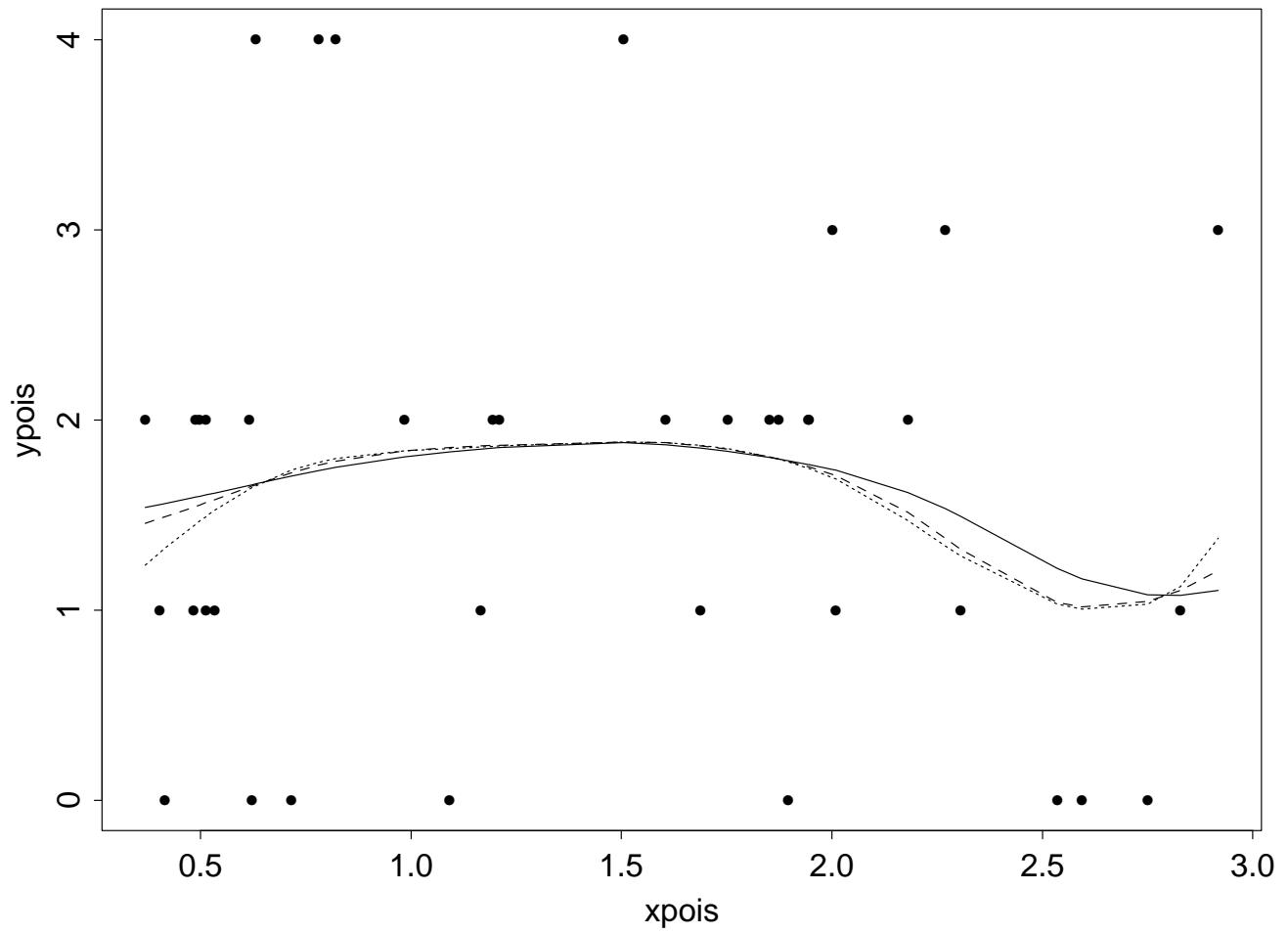
$$\log\{\bar{Y}_w\} + (x - \bar{X}_w) \frac{\sigma_w(X, Y)}{h^2 \mu_2 \bar{Y}_w} + \frac{\sigma_w(X, Y)^2}{2h^2 \mu_2 \bar{Y}_w^2}$$

- Logistic Regression:

$$\text{logit}\{\bar{Y}_w\} + (x - \bar{X}_w) \frac{\sigma_w(X, Y)}{h^2 \mu_2 \bar{Y}_w (1 - \bar{Y}_w)} + \frac{(1 - 2\bar{Y}_w) \sigma_w(X, Y)^2}{2h^2 \mu_2 \bar{Y}_w^2 (1 - \bar{Y}_w)^2}$$

- Inherits boundary corrections: $\tilde{\mathbf{a}} - \hat{\mathbf{a}}$





Summary

- Likelihood Cross-Validation and h
- Relationship to methods for measurement error
- Symbolic Computation vs Boundary kernels