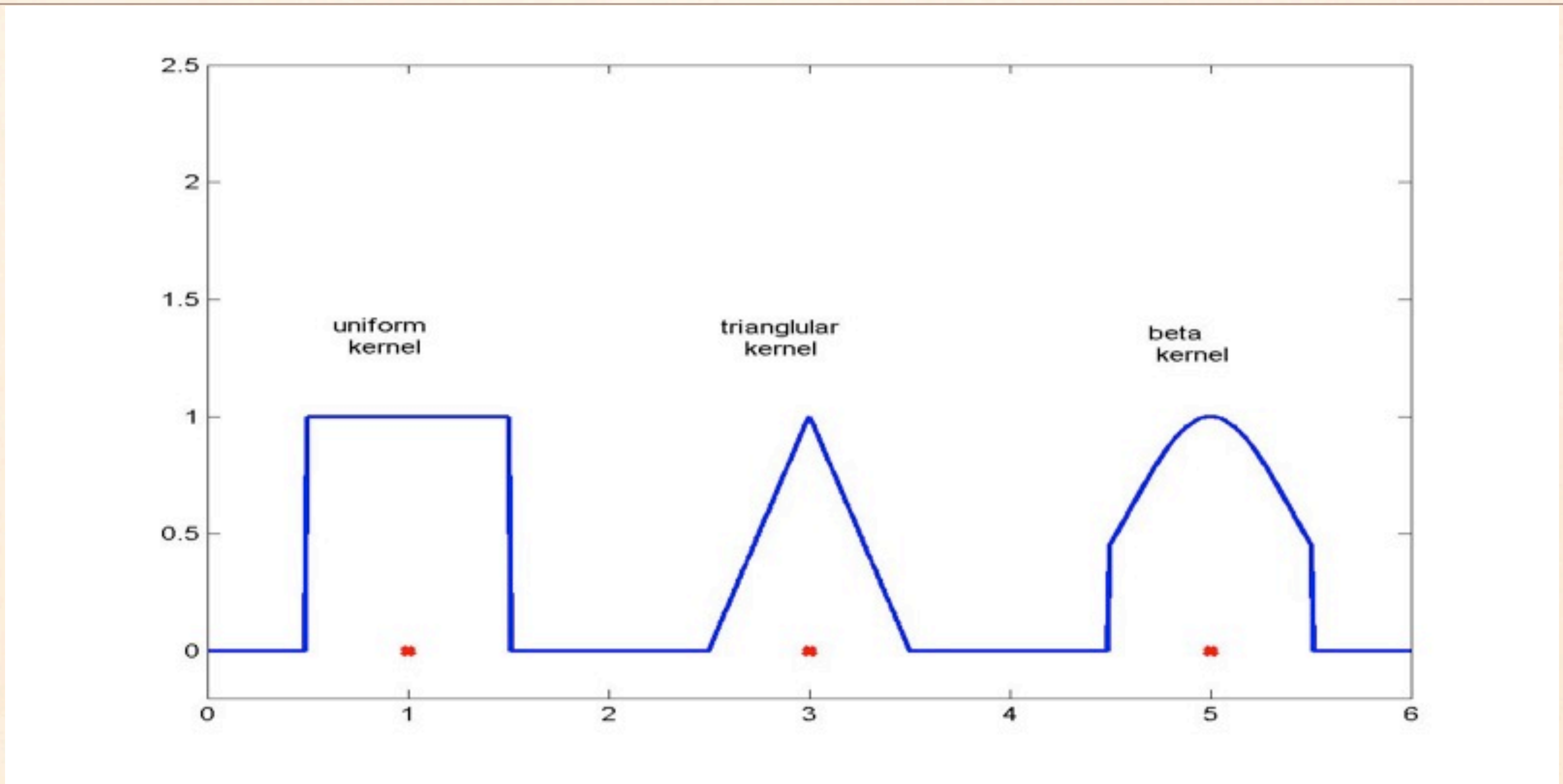
OPTIMIZATION TECHNOLOGY TO DEAL WITH LACK OF DATA IN STATISTICAL ESTIMATION

Roger J-B Wets

Mathematics, University of California, Davis

Wisconsin, Spring 2011

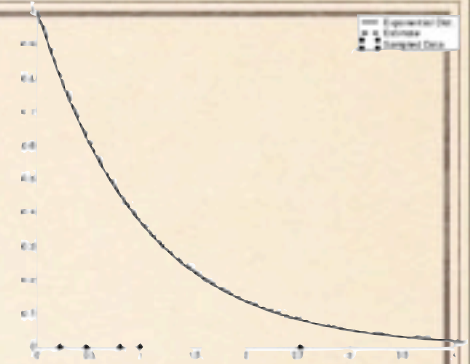
Kernel "Base" NonParametric Estimation



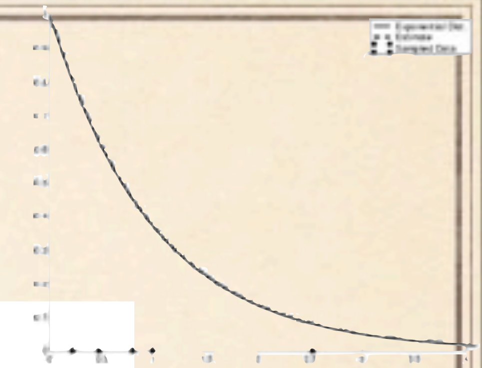
Information = Observations: $\xi^1, \xi^2 \ldots, \xi^l$

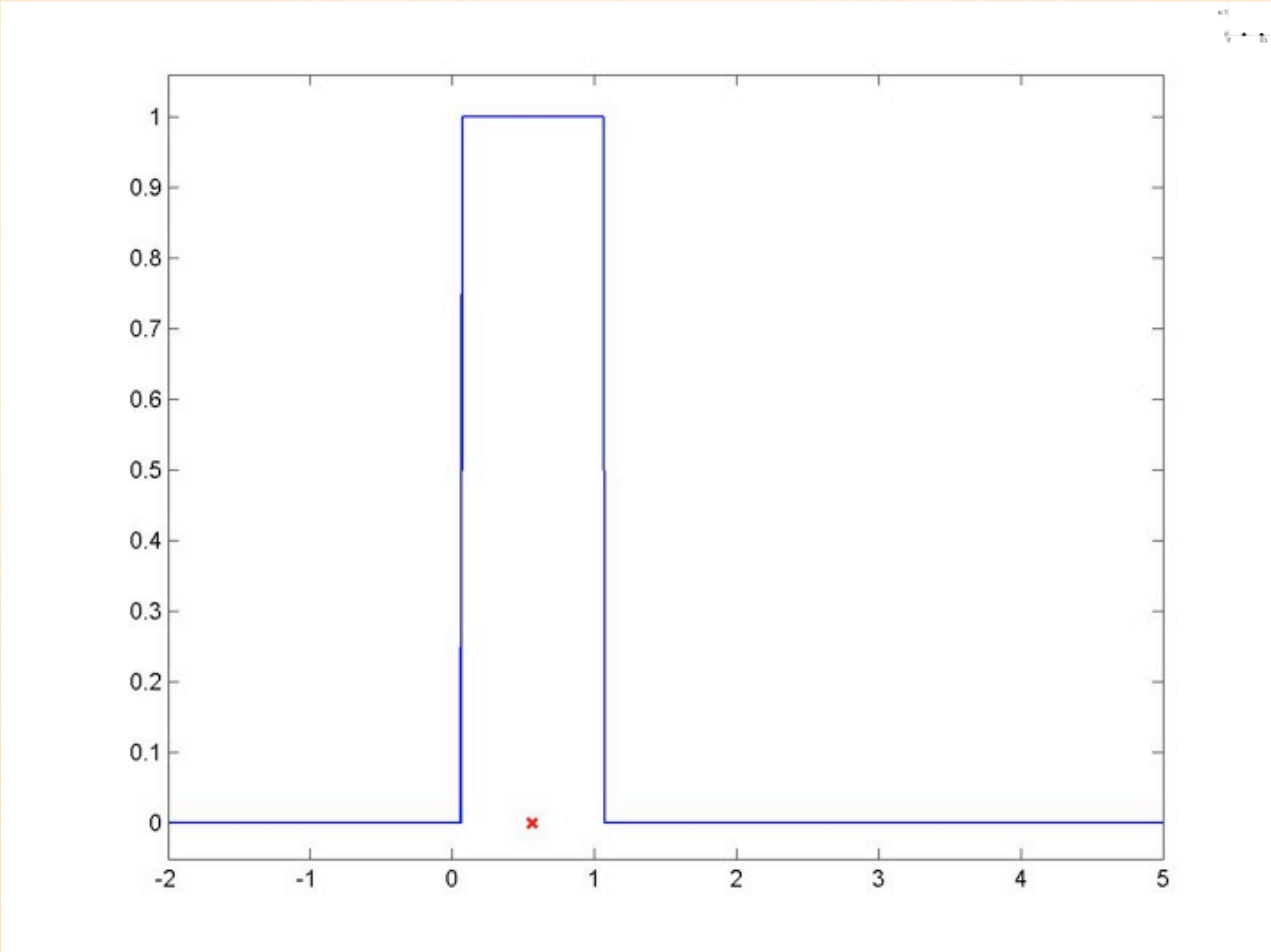
Optimal bandwidth = kernel support ?

Kernel Estimates

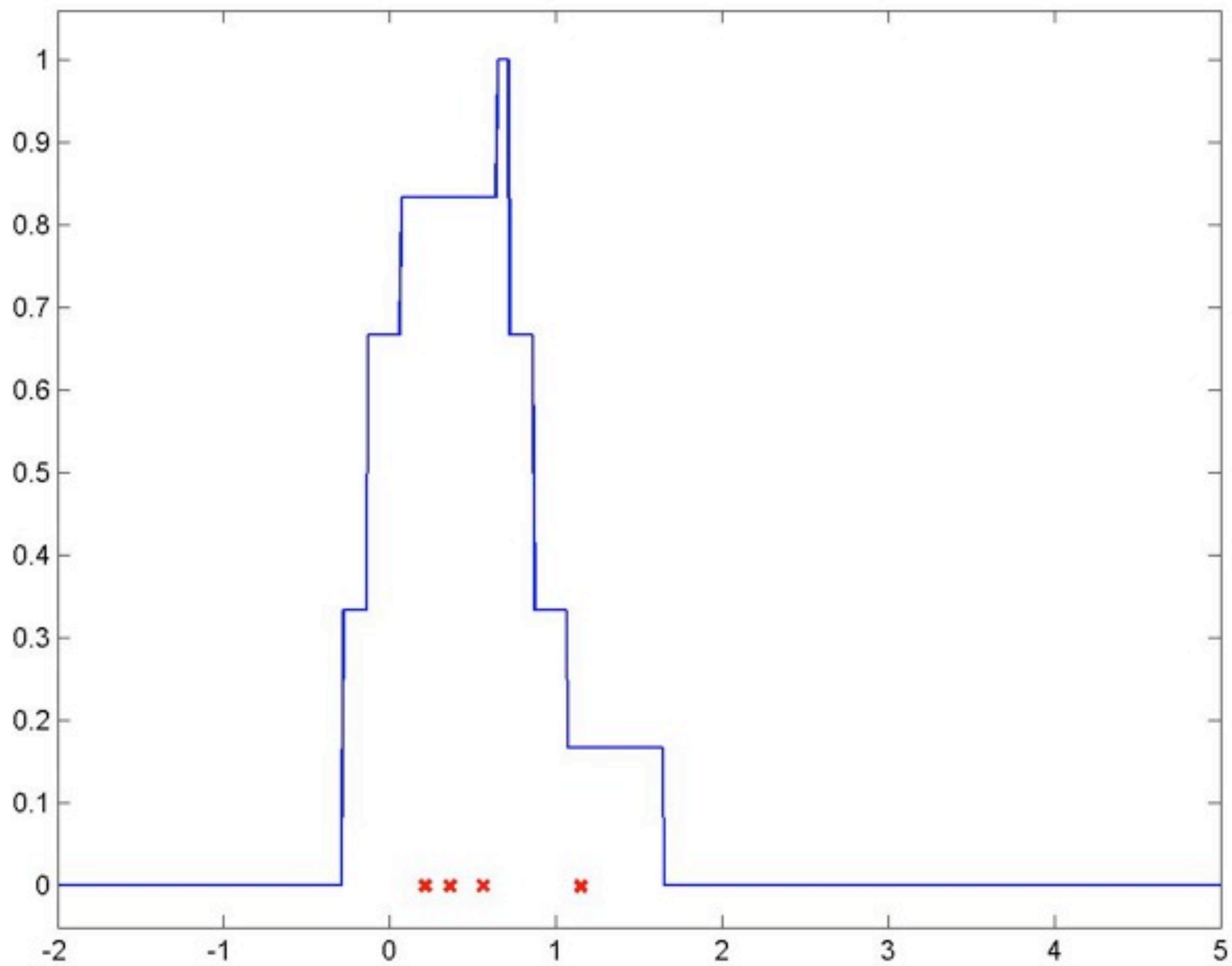


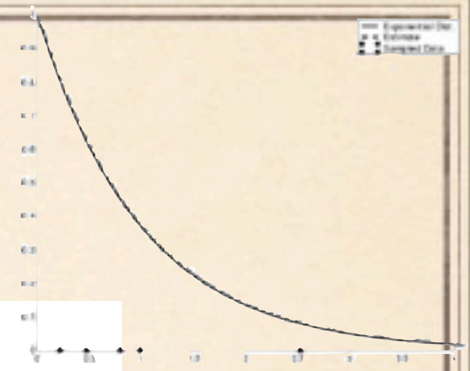
Kernel Estimates



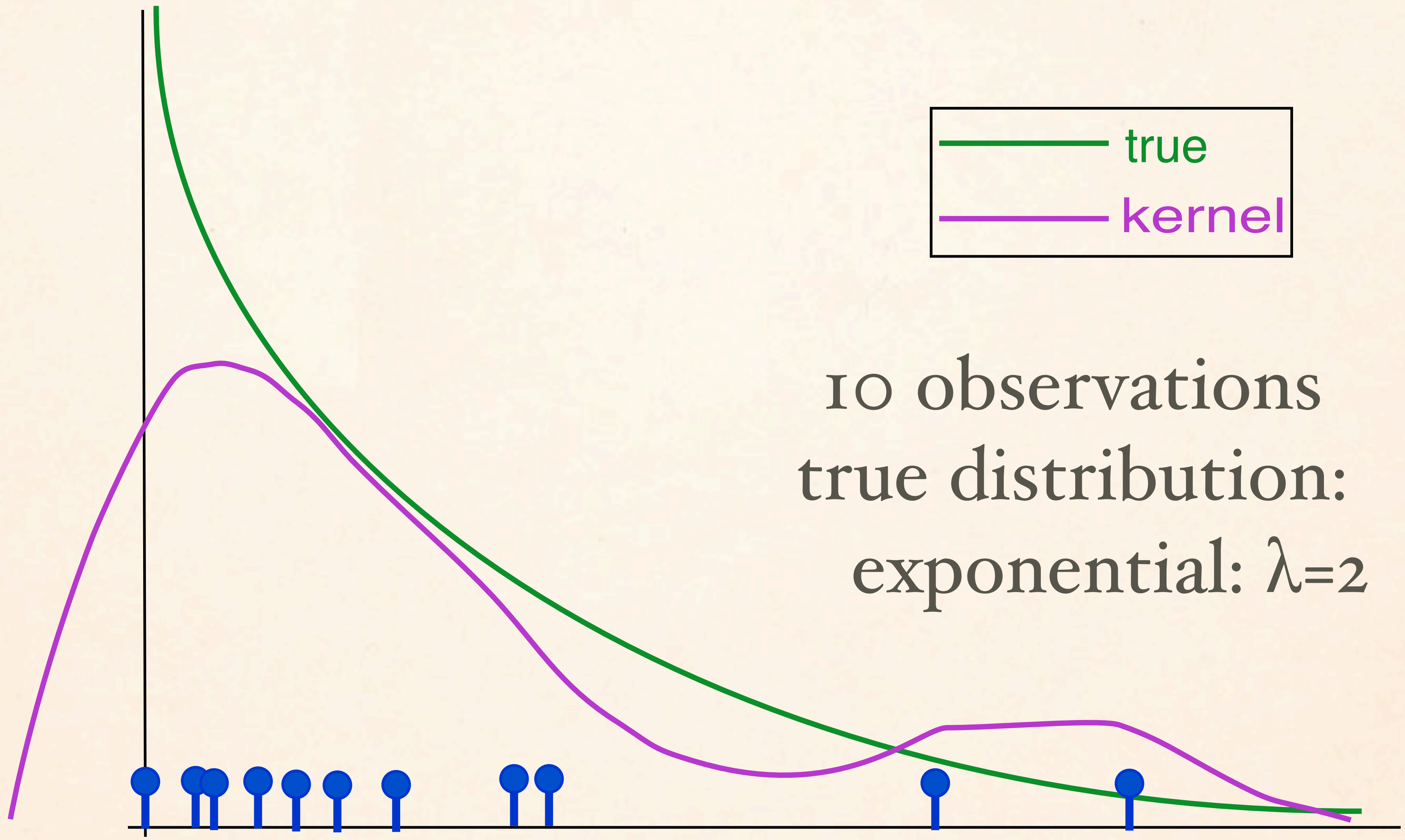


Kernel Estimates



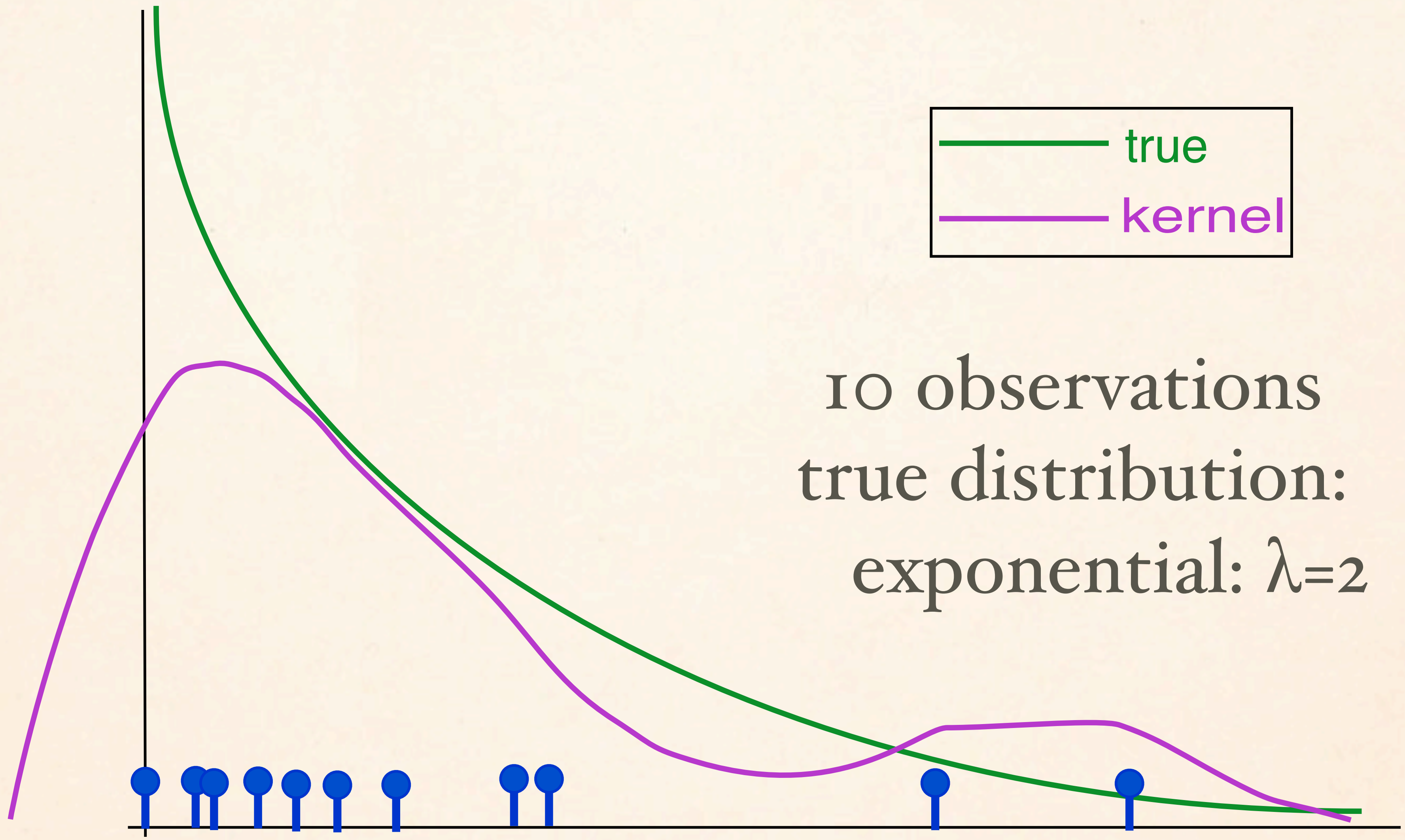


R-STAT: KERNEL ESTIMATE



10 observations
true distribution:
exponential: $\lambda=2$

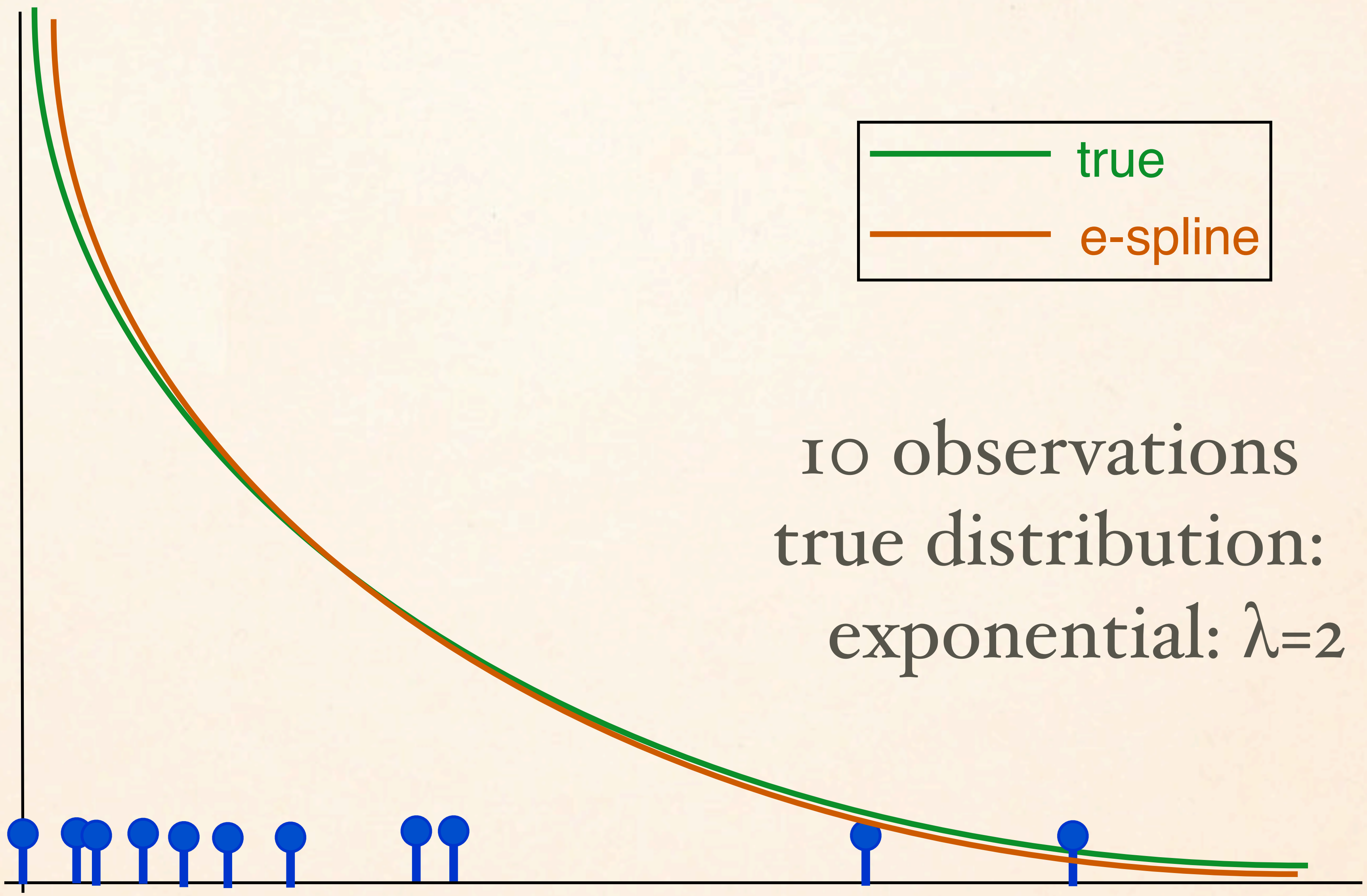
R-STAT: KERNEL ESTIMATE



10 observations
true distribution:
exponential: λ=2

Maybe an 'optimizer's' viewpoint might help?

"OUR" ESTIMATE



10 observations
true distribution:
exponential: λ=2

BASIC STATISTICAL ESTIMATION PROBLEM

- Find F^{est}, an estimate of the distribution of ξ given **all** the information available about this random phenomena,
- i.e. such that

$$\forall z: \ F^{est}(z) \approx F^{true}(z) = \text{prob.}[\xi \le z]$$

all information might come with inequalities ....

“ALL” INFORMATION

- Observations (hard data): $\xi^1, \xi^2, \ldots, \xi^\nu$
- Non-data facts (soft information)
  - Support: (un)bounded , density or discrete distribution,
  - bounds on expectation, moments,
  - heavy tails
  - shape: unimodal, decreasing, parametric class
- still `softer' information (modeling assumptions):
  - see above + ... level of smoothness, `Bayesian' neighborhood, ..

AN OPTIMIZATION VIEWPOINT

- Find *h in H=class-fcns*$(\mathbb{R}^n)$
- that maximizes the probability of observing $\xi^1, \xi^2, \ldots, \xi^\nu$

$$\max \boxed{\frac{1}{\nu}\sum_{l=1}^{\nu} \ln h(\xi^l) \quad \text{(likelihood)}}$$

- $$\max E^\nu \left\{ \ln h(\xi) \right\} = \max \int \ln h(\xi) P^\nu(d\xi)$$

"SOFT" INFORMATION

"SOFT" INFORMATION

support: $S = \left[\alpha, \beta\right],\ S = [\alpha, \infty), \ldots$

"SOFT" INFORMATION

support: $S = [\alpha, \beta],\ S = [\alpha, \infty), \ldots$

bounds on moments: $a \le \int \xi h(\xi)\, d\xi \le b$

"SOFT" INFORMATION

support: $S = [\alpha, \beta],\ S = [\alpha, \infty), \ldots$

bounds on moments: $a \le \int \xi h(\xi)\, d\xi \le b$

unimodal: $h(x) = e^{-Q(x)}$, Q convex (Hessian SDP)

“SOFT” INFORMATION

support: $S = [\alpha, \beta],\ S = [\alpha, \infty), \ldots$

bounds on moments: $a \leq \int \xi h(\xi) d\xi \leq b$

unimodal: $h(x) = e^{-Q(x)}$, Q convex (Hessian SDP)

decreasing: $h'(x) \leq 0,\ \forall x \in S$

"SOFT" INFORMATION

support: $S = [\alpha, \beta],\ S = [\alpha, \infty), \ldots$

bounds on moments: $a \le \int \xi h(\xi)\, d\xi \le b$

unimodal: $h(x) = e^{-Q(x)}, Q$ convex (Hessian SDP)

decreasing: $h'(x) \le 0,\ \forall x \in S$

'smoothness': $h \in H_0^1,\ \int \frac{h'(\xi)^2}{h(\xi)} d\xi \le \kappa$ (Fisher info.)

"SOFT" INFORMATION

support: $S = [\alpha, \beta],\ S = [\alpha, \infty), \ldots$

bounds on moments: $a \le \int \xi h(\xi)\, d\xi \le b$

unimodal: $h(x) = e^{-Q(x)}$, Q convex (Hessian SDP)

decreasing: $h'(x) \le 0,\ \forall x \in S$

'smoothness': $h \in H_0^1,\ \int \frac{h'(\xi)^2}{h(\xi)} d\xi \le \kappa$ (Fisher info.)

'Bayesian': $\|h - h^0\| \le \beta$, objective: $\alpha \cdot \text{bayes}(h, h^0)$

RE-FORMULATION: "OPT"-VERSION

$$\max E^\nu\left\{\ln h(x)\right\} = \frac{1}{\nu}\sum_{l=1}^{\nu} \ln h(x_l)$$

$$\text{such that } \int h(x)dx = 1,$$

$$h(x) \geq 0,\ \forall x \in \mathbb{R}$$

$$h \in A^\nu \subset H$$

A^ν: soft (non-data) information constraints

$H = C^2(S), L^p(S), H^1(S), \ldots \quad S$ subset $\mathbb{R}^n$

CONSISTENCY THEOREM?

Suppose $\nu \to \infty$ (more data is acquired) and

$A^\nu \to A$ (valid information is acquired) then

estimates $h^\nu \to h^{true}$ $a.s.$ (with probability 1).

1949 A. Wald: consistency of parametric estimates (MLH-SAA)

1982-1983 Klonias & Prakasa Rao: consistency of nonparametric estimatesr

1971 Good & Gaskins: nonparametric roughness penalties (proposed)

1982 B. Silverman & 1990 J. Thompson/R.Tapia: consistency with penalization

1985 P. Groeneboom: Estimating a monotone density

==

1979 R. Wets: statistical approach to solution of stochastic program (Tech. Note)

1988 (with J. Dupacova): asymptotics of constrained estimators (parametric)

1991 (with H. Attouch): Law of Large Numbers for random lsc functions

2000 (with X. Dong): consitency of constrained estimators (non-parametric)

2006 (with M. Casey): rates of convergence

APPROXIMATION THEORY

$\min f_0(x)$ such that $x \in X \subset H$ (for our use: H Polish space)

$\min f(x)$ on $x \in X$ with $f(x) = \begin{cases} f_0(x) & \text{when } x \in X \\ \infty & \text{otherwise} \end{cases}$ lsc function: $H \to \overline{\mathbb{R}}$

$(f_0^\nu, X_0^\nu) \to (f_0, X)$ sequence of optimization problems converging(?) to, f

$\arg\min f^\nu \to \arg\min f \quad (\inf f^\nu \to \inf f)$

$\arg\min\,(f^\nu + g) \to \arg\min(f + g),\ g$ continuous perturbation

$\Rightarrow f^\nu$ epi-converges to f $\quad (\operatorname{epi} f^\nu \to \operatorname{epi} f)$: for all $x \in H$,

(a) $\forall x^\nu \to x,\ \liminf_\nu f^\nu(x^\nu) \geq f(x)$,

(b) $\exists x^\nu \to x,\ \limsup_\nu f^\nu(x^\nu) \leq f(x)$.

pointwise?, uniform?,

LLN: RANDOM LSC FUNTIONS

$f:\Xi\times H\to\overline{\mathbb{R}}$ a random lsc function, ξ values in $(\Xi,\mathcal{A},P)$

(a) lsc (lower semicontinuous) in h, $(\forall\xi\in\Xi)$

(b) (ξ,h)-measurable $\qquad (\mathcal{A}\times B_X)$-measurable

recall: $f(\xi,h)= f_0(\xi,h)$ when $h\in X(\xi)$ -- stochastic constraints

$$f^\nu(\xi,h)=\begin{cases}\frac{1}{\nu}\sum_{l=1}^{\nu}\ln h(\xi^l) \text{ if } h\geq 0,\ \int_\Xi h(\xi)d\xi=1,\ h\in A^\nu\\ \infty \text{ otherwise} \quad (\sim \text{SAA of optimisation problems})\end{cases}$$

Question: Do the $f^\nu(\xi,\cdot)$ epi-converge to $\mathbb{E}\{f(\xi,h)\}$ P-a.s.?

$$h^{true}\in\arg\min\ \mathbb{E}\{f(\xi,h)\}$$

where $\quad f(\xi,h)=\begin{cases}\ln h(\xi) \text{ if } h\geq 0,\ \int_\Xi h(\xi)d\xi=1, h\in A\\ \infty \text{ otherwise}\end{cases}$

CONSISTENCY THEOREM

Suppose $\nu \to \infty$ (more data is acquired) and

$A^{\nu} \to A$ (valid information is acquired) then

estimates $h^{\nu} \to h^{true}$ $a.s.$ (with probability 1).

NUMERICAL STRATEGIES

$$h \approx \sum_{k=1}^{q} u_k \phi_k(\cdot)$$

Fourier coefficients, wavelets, kernel-dictionary, …

$$h = \exp\big(s(\bullet)\big) \quad \text{exponential epi-spline}$$

$s(\bullet)$ cubic (or quadratic) epi-spline, spline-like

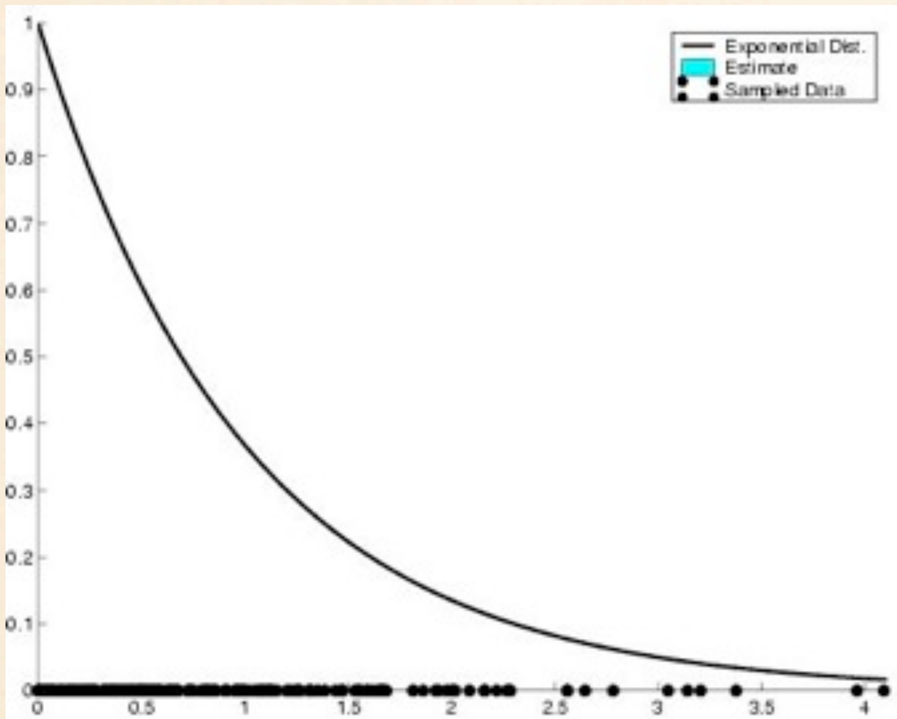
$\Rightarrow$ n-dimensional theory of epi-splines

TEST CASE: EXPONENTIAL

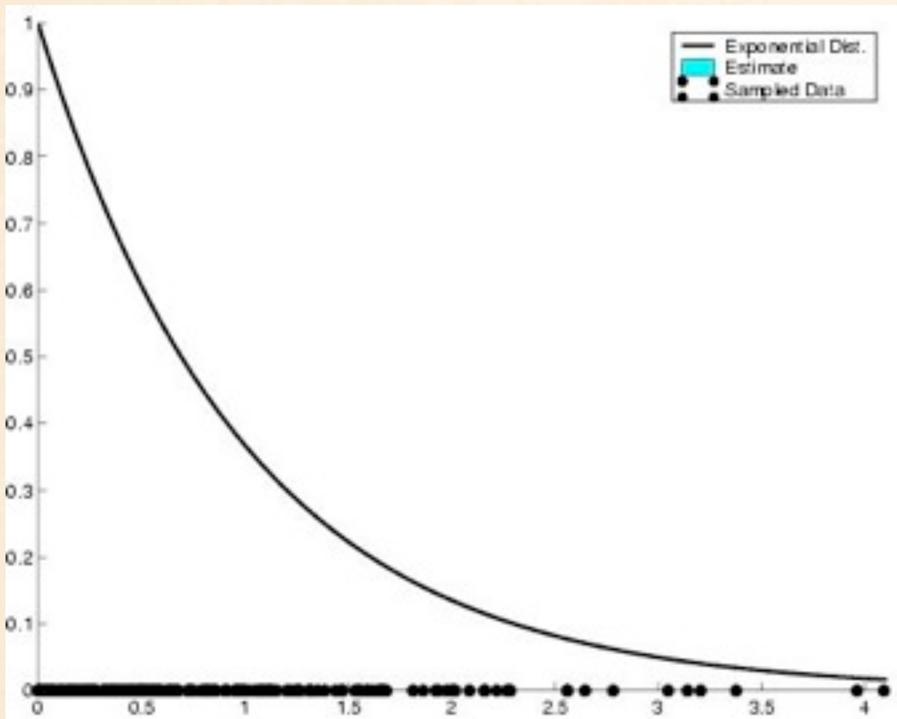
- $h^{true}(x) = \lambda e^{-\lambda x}$ if $x \geq 0$; $= 0$ if $x < 0$ $(\lambda = 1)$
- "empirical" estimate
- kernel estimate from **R-stat**
- unconstrained with support (non-negative)
- constrained (*h* decreasing)
- parametric, i.e., $h \in \exp\text{-class}$

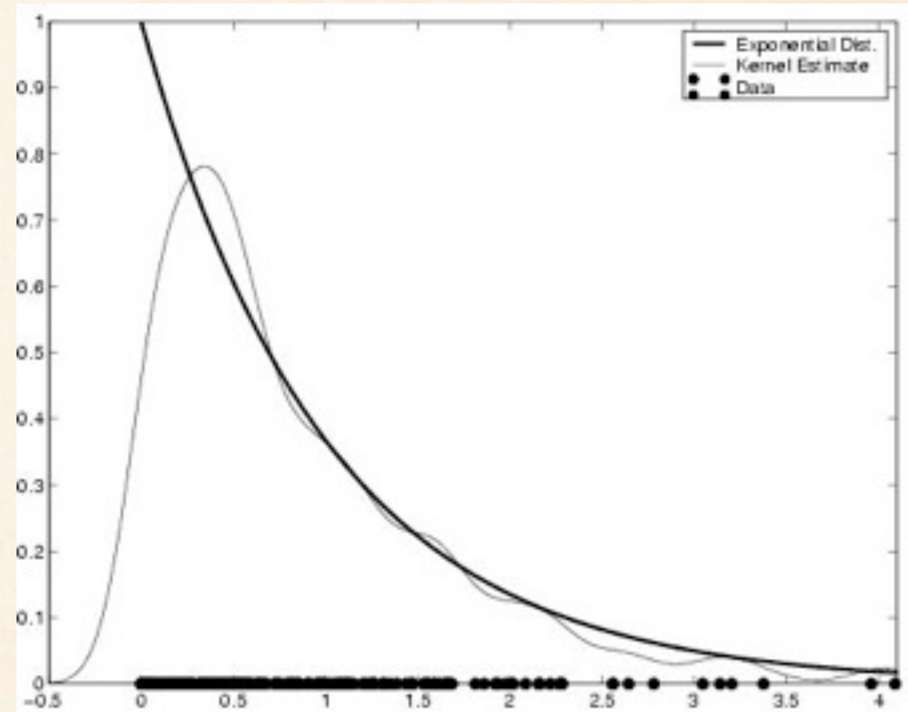
200-OBSERVATIONS

200-OBSERVATIONS

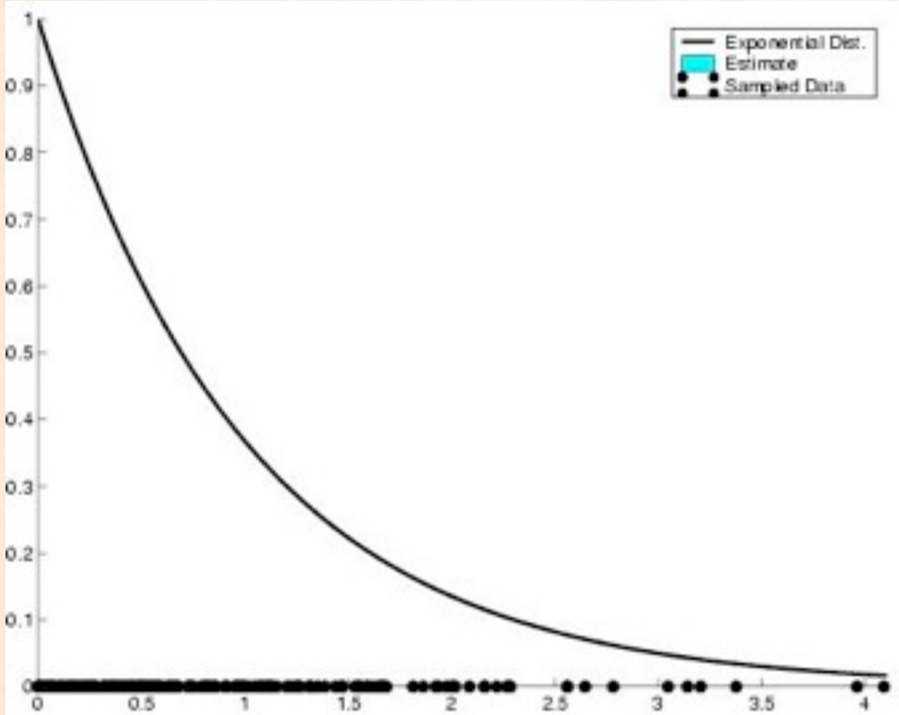


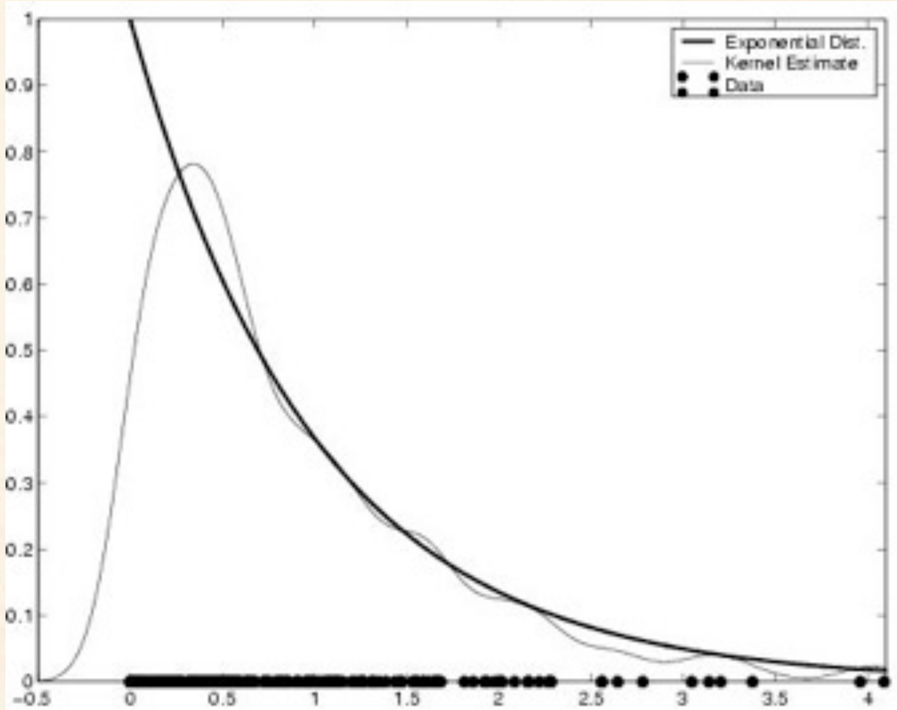
200-OBSERVATIONS

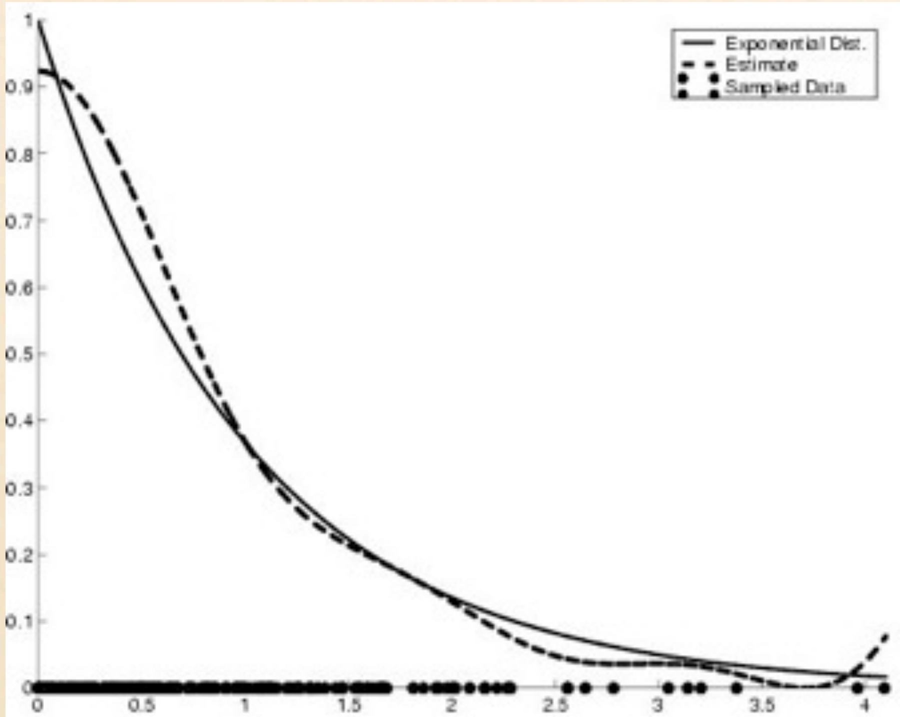




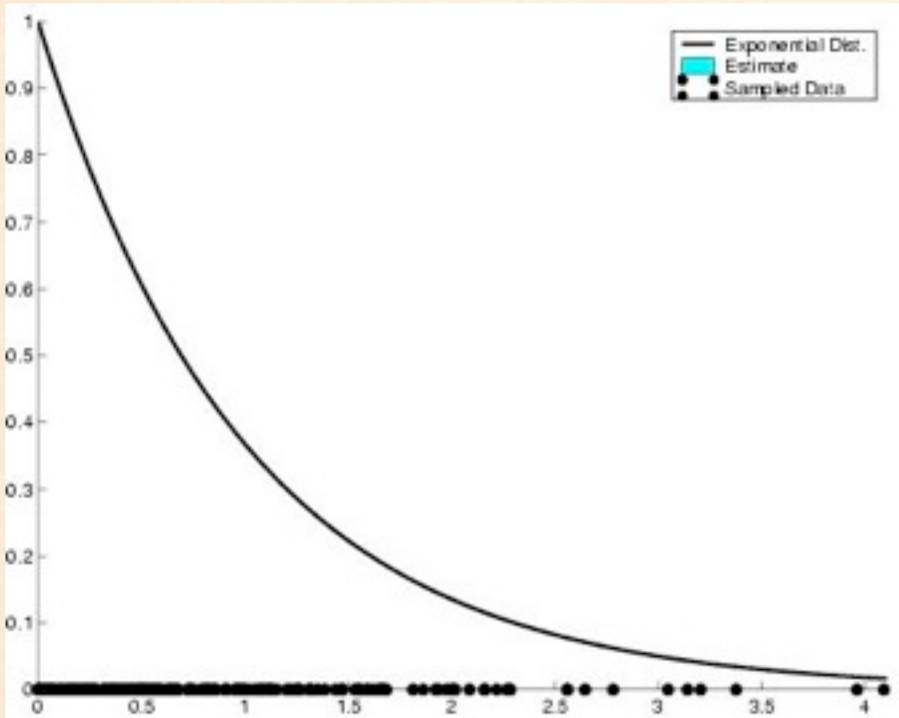
200-OBSERVATIONS

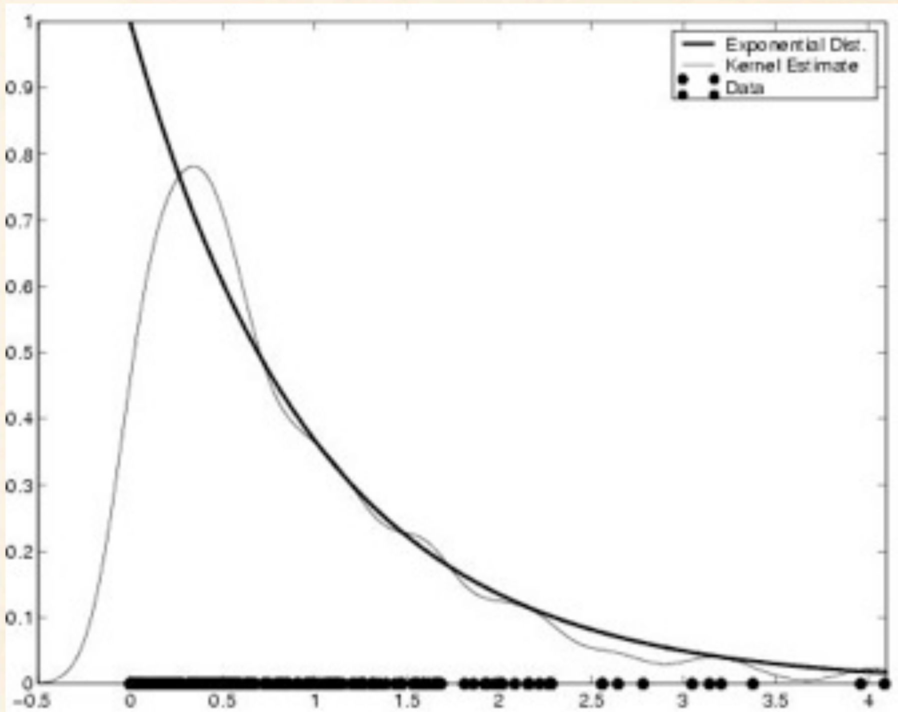


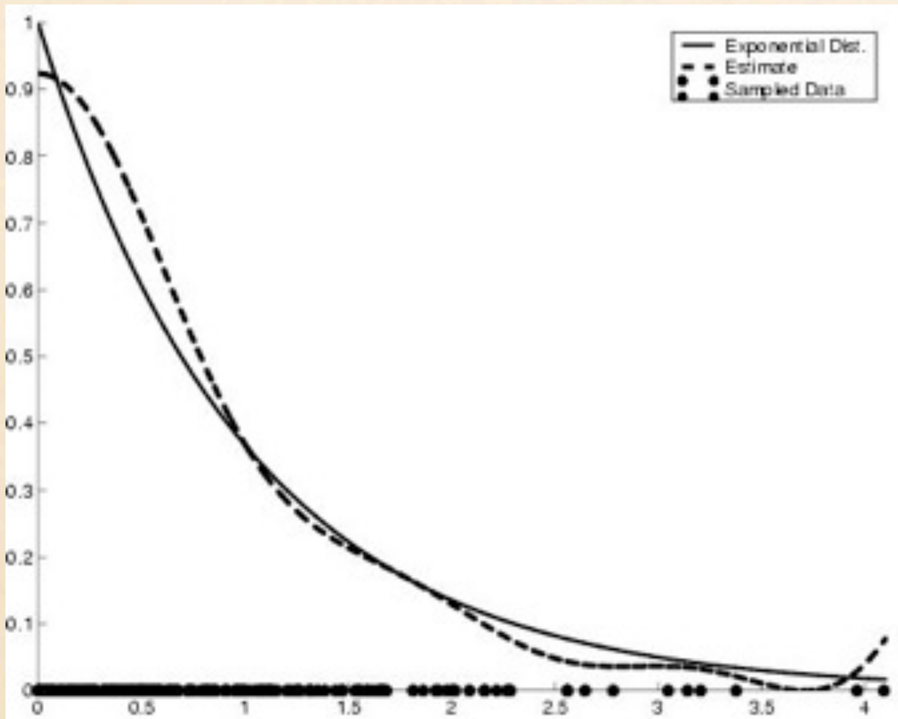


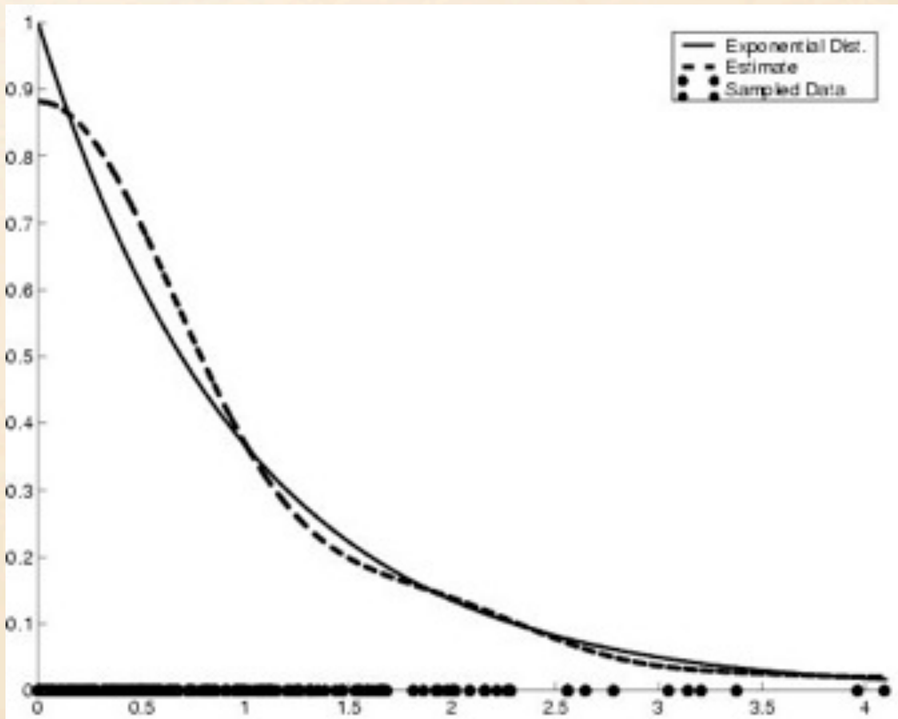


200-OBSERVATIONS

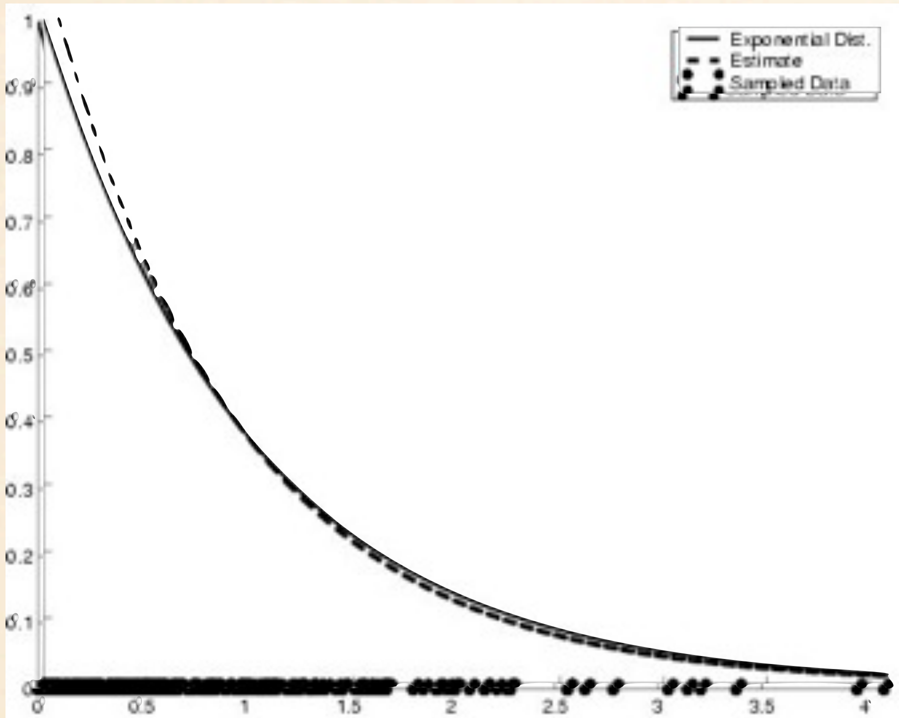


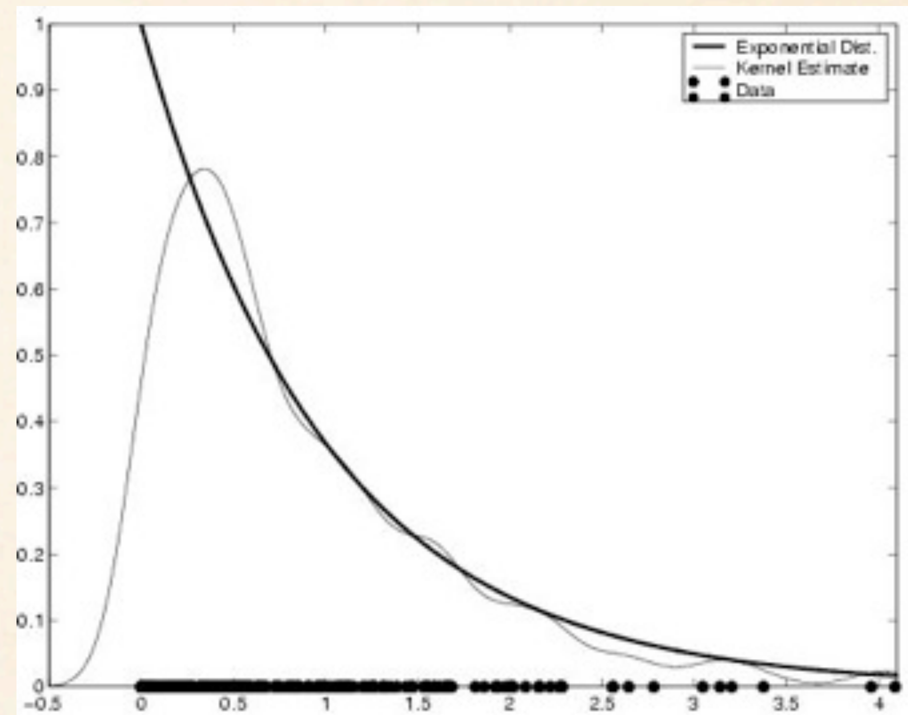


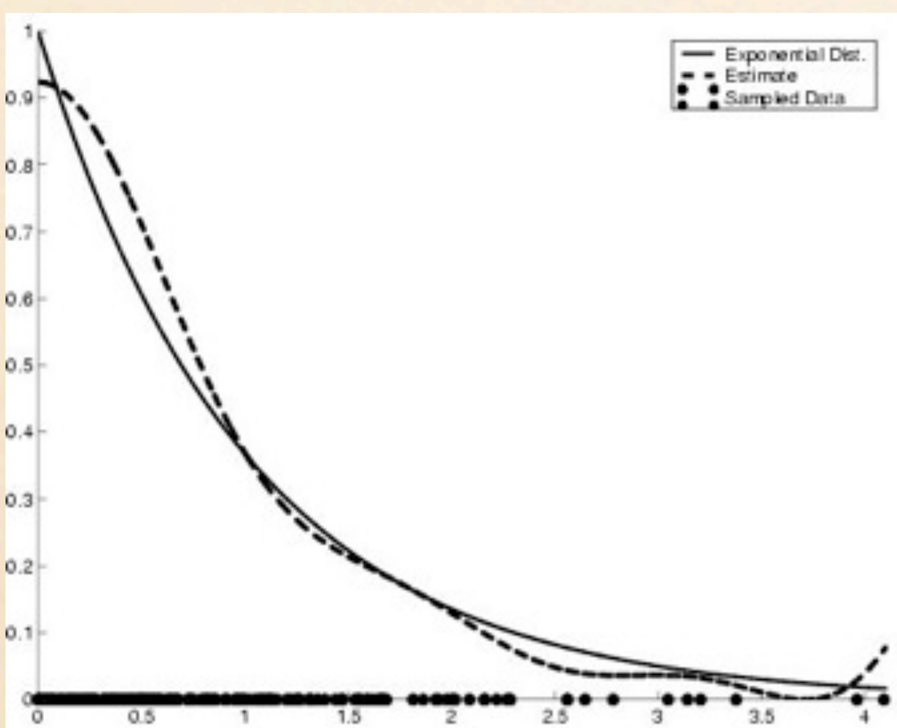


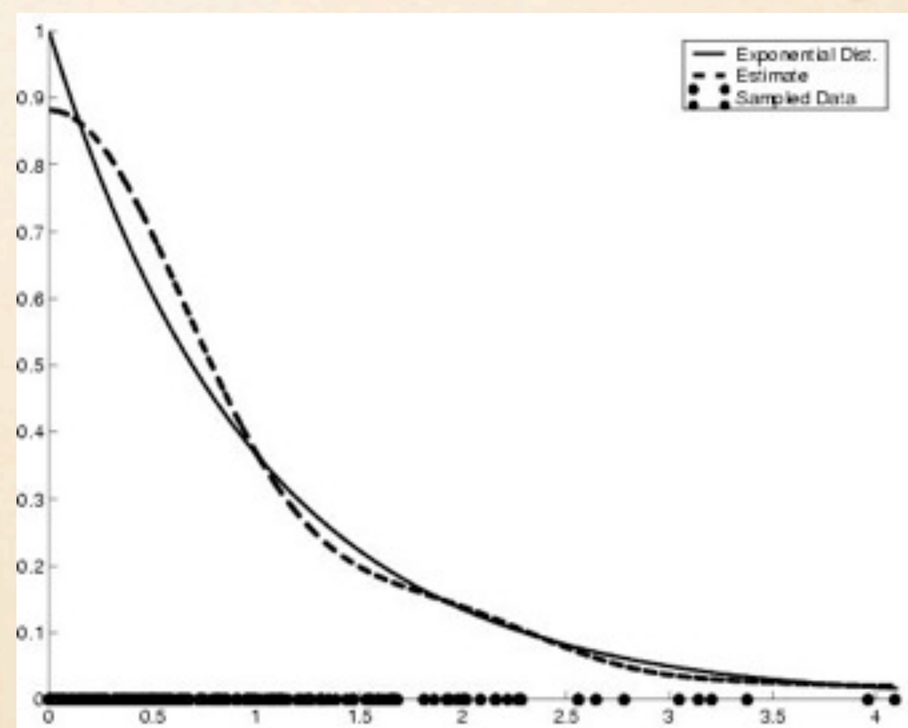


200-OBSERVATIONS



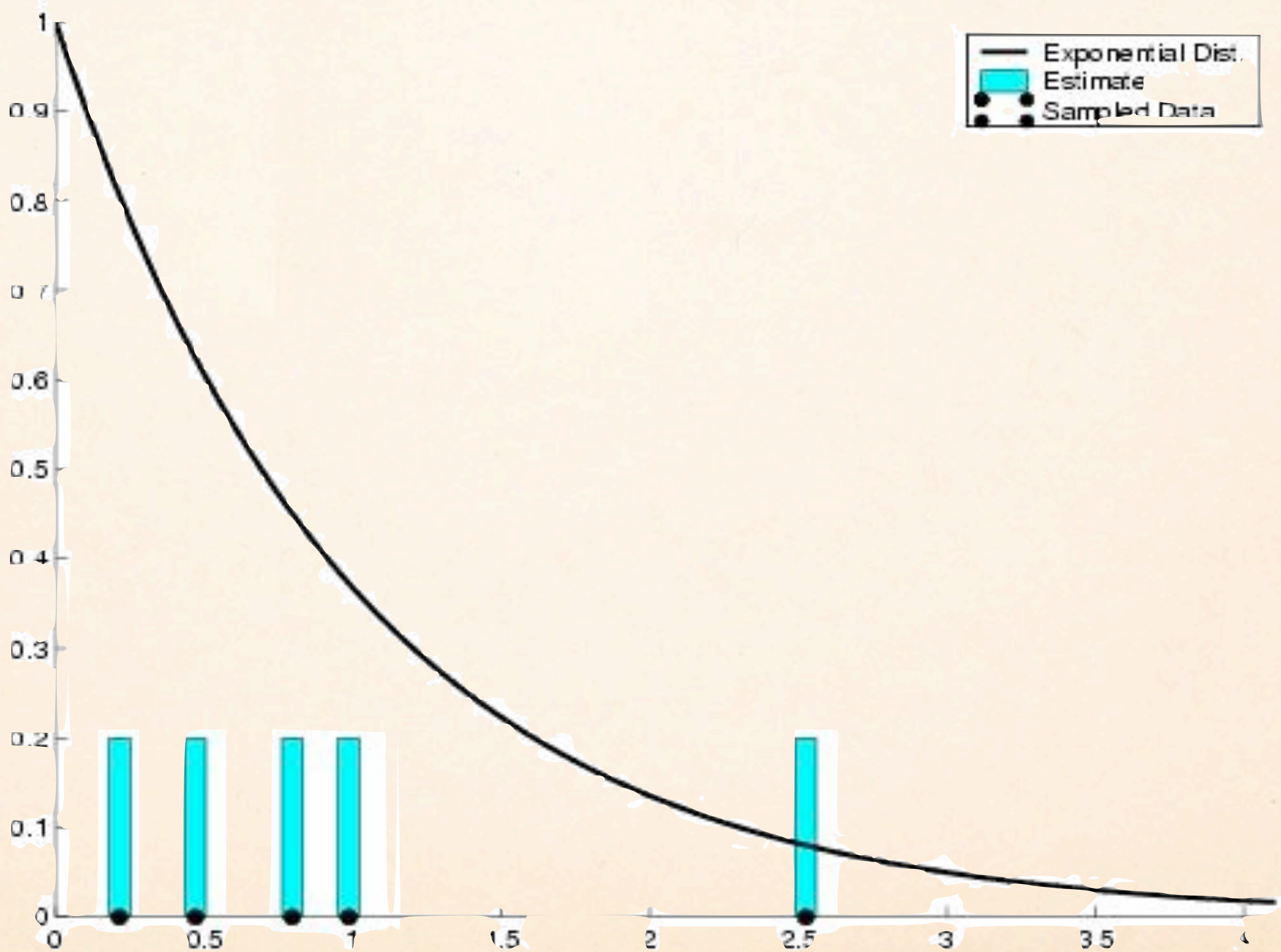




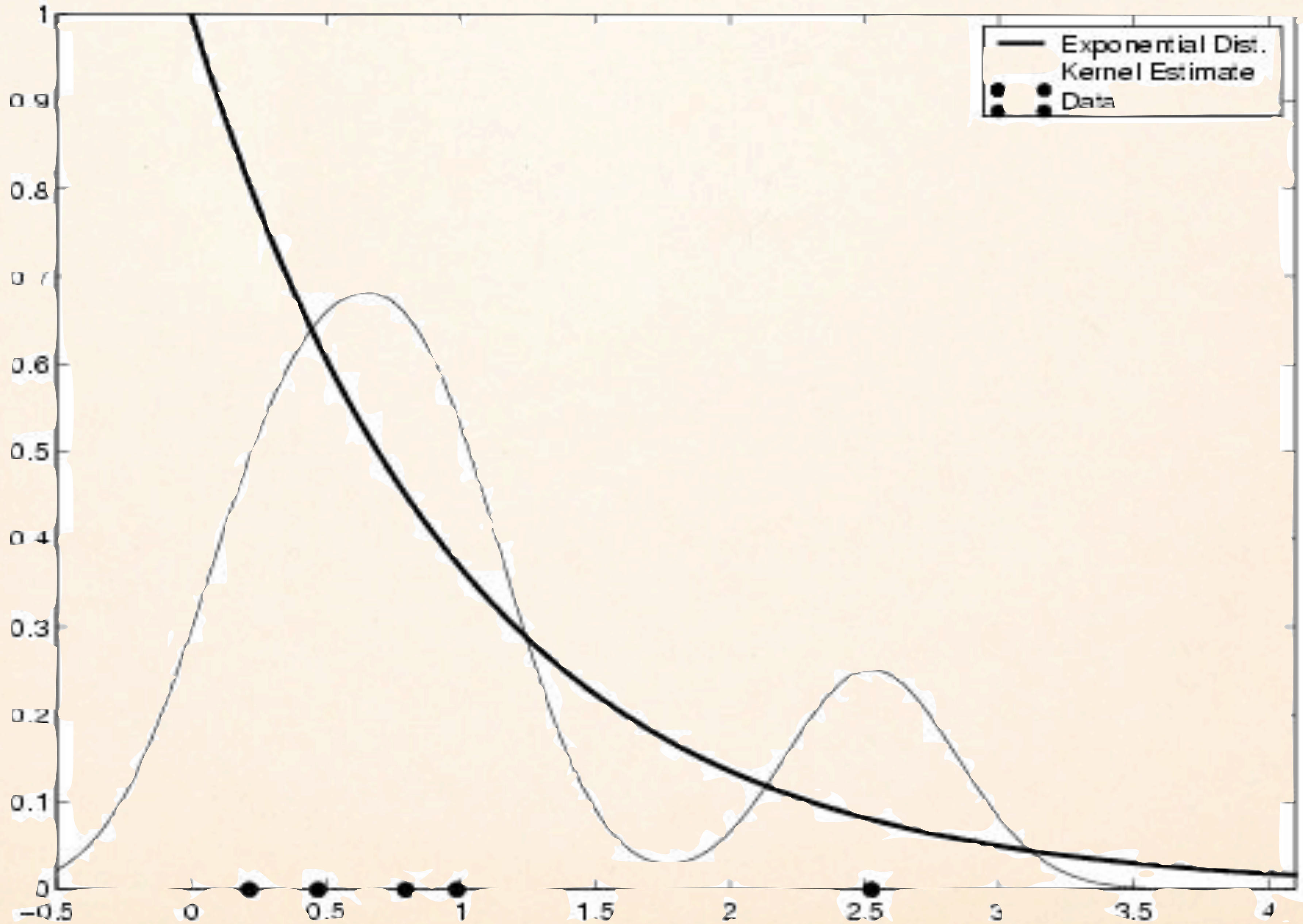


But what about 5 sample points?

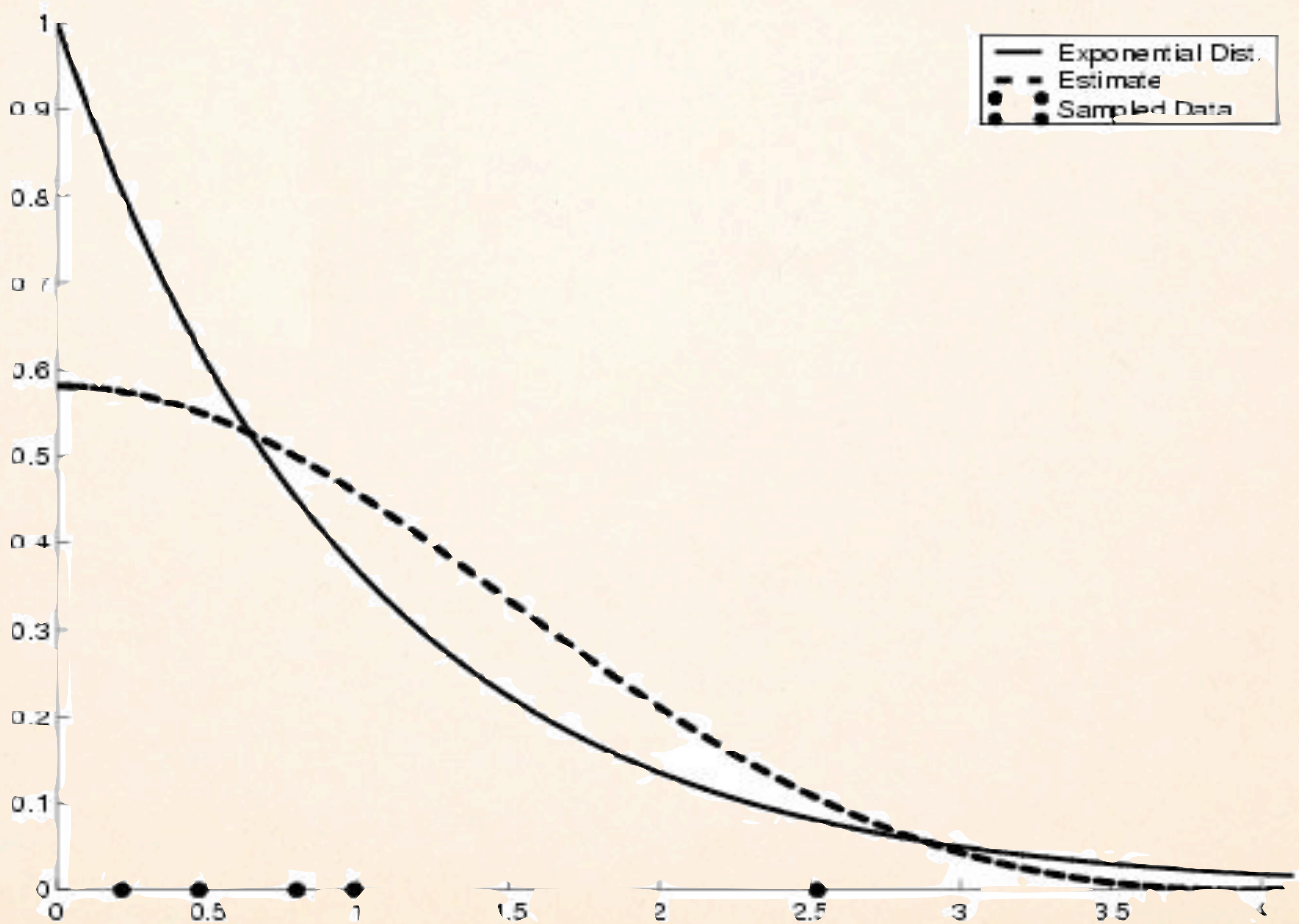
h^{est} : EMPIRICAL ESTIMATE



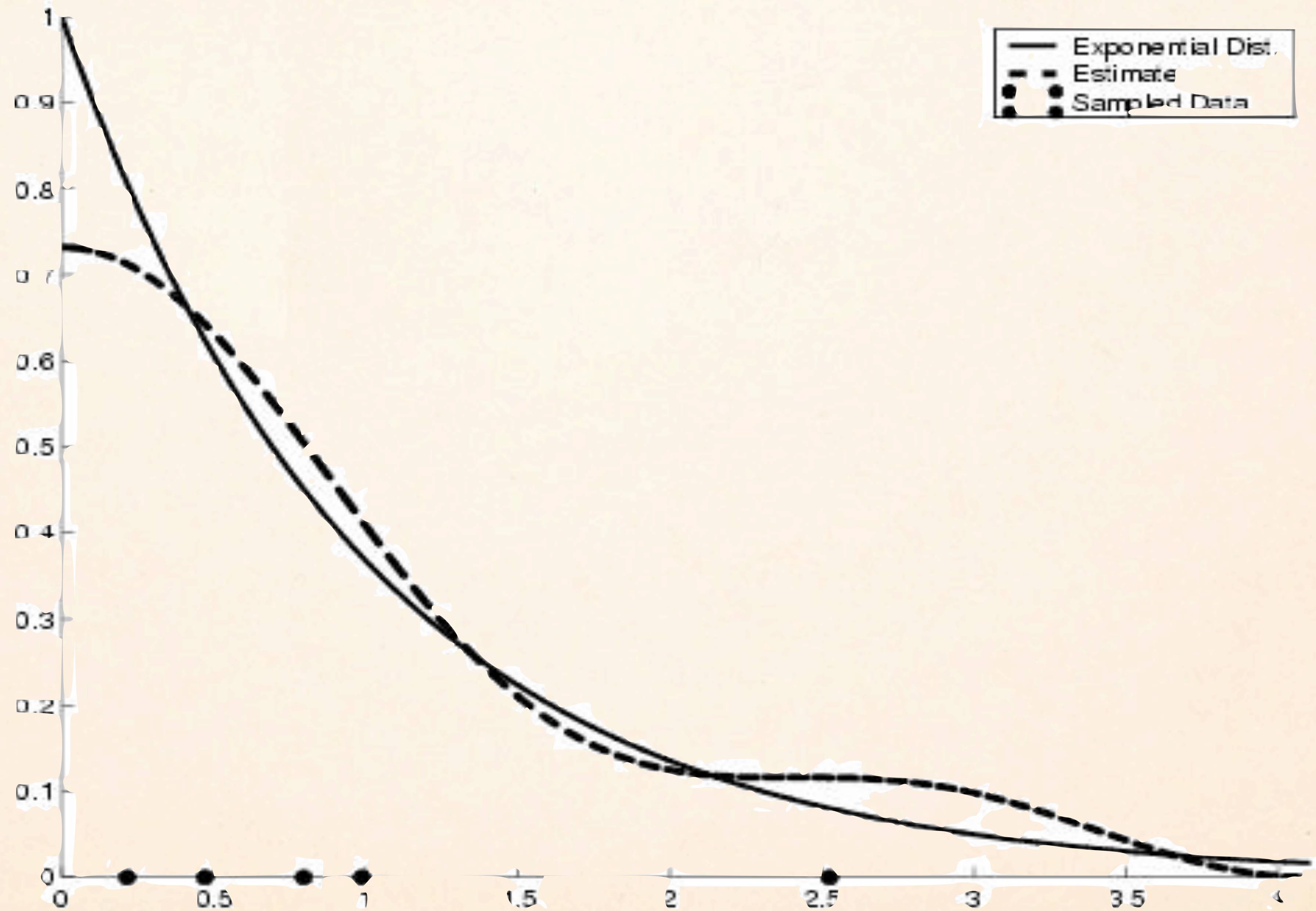
h^{est}: KERNEL ESTIMATE (R-STAT)



h^{est}: ..WITH KNOWN SUPPORT

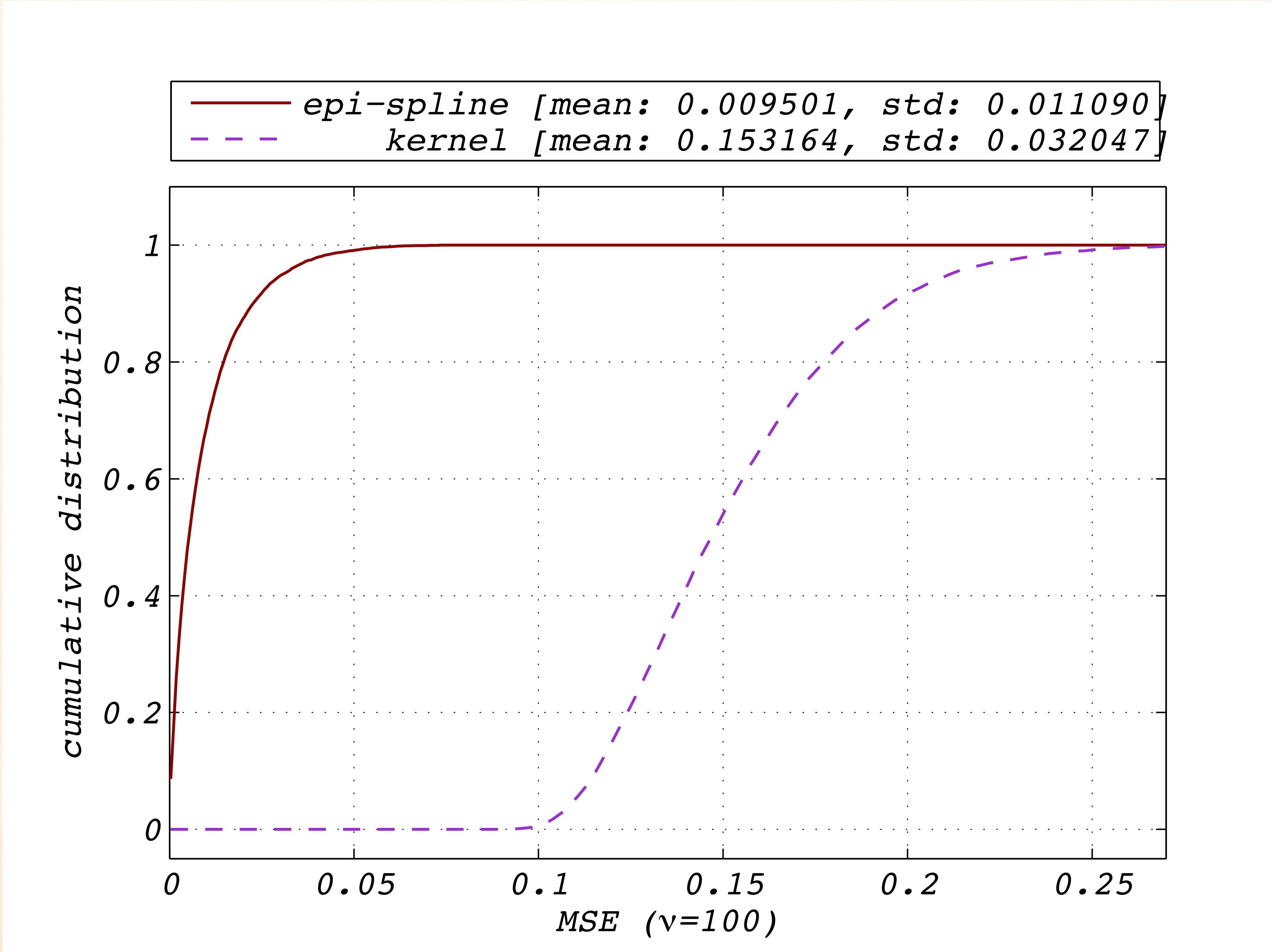


h^{est} : DECREASING DENSITY



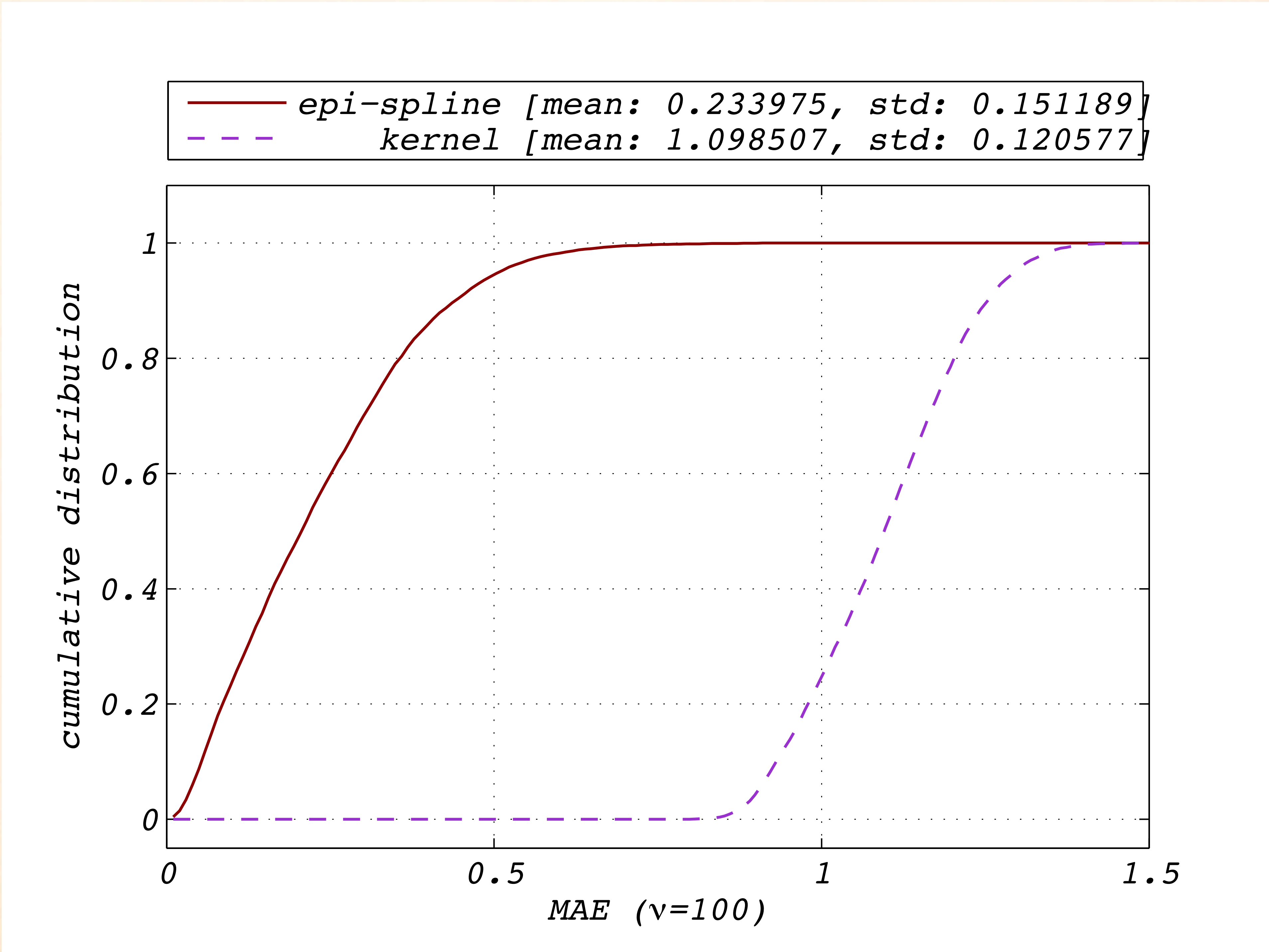
DOES IT PAY OFF? ERROR ANALYSIS

DOES IT PAY OFF? ERROR ANALYSIS

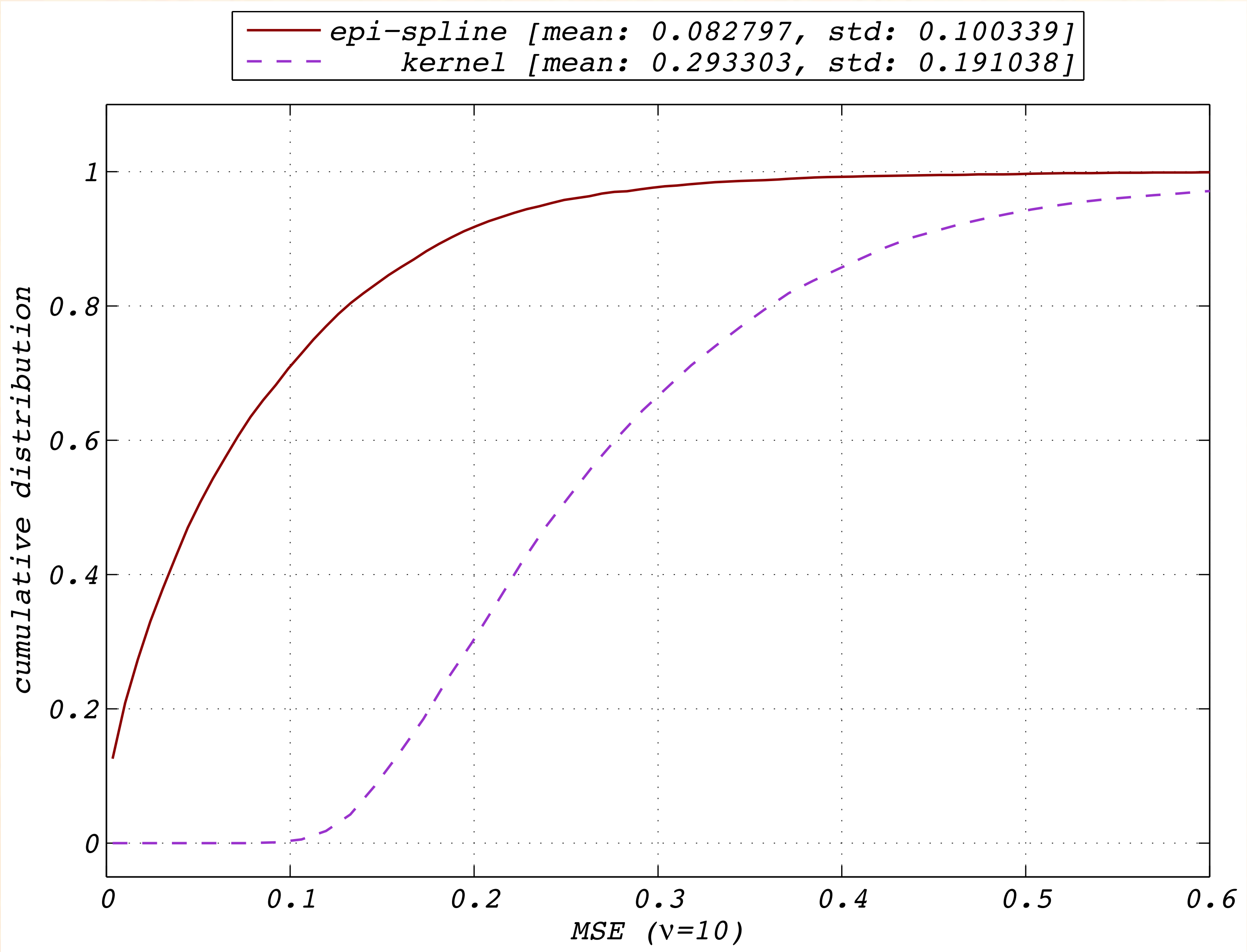


DOES IT PAY OFF? ERROR ANALYSIS

DOES IT PAY OFF? ERROR ANALYSIS

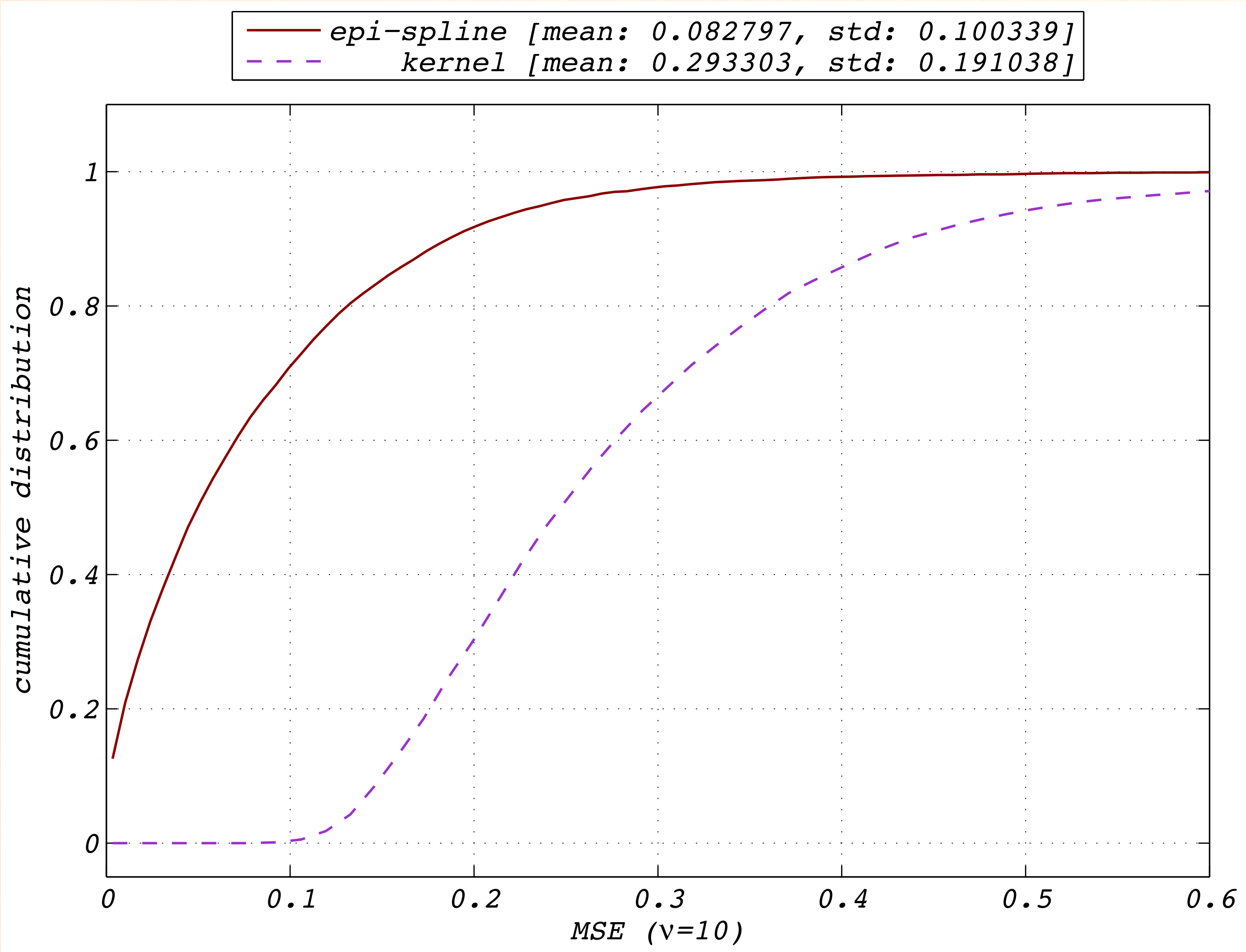


DOES IT PAY OFF? ERROR ANALYSIS

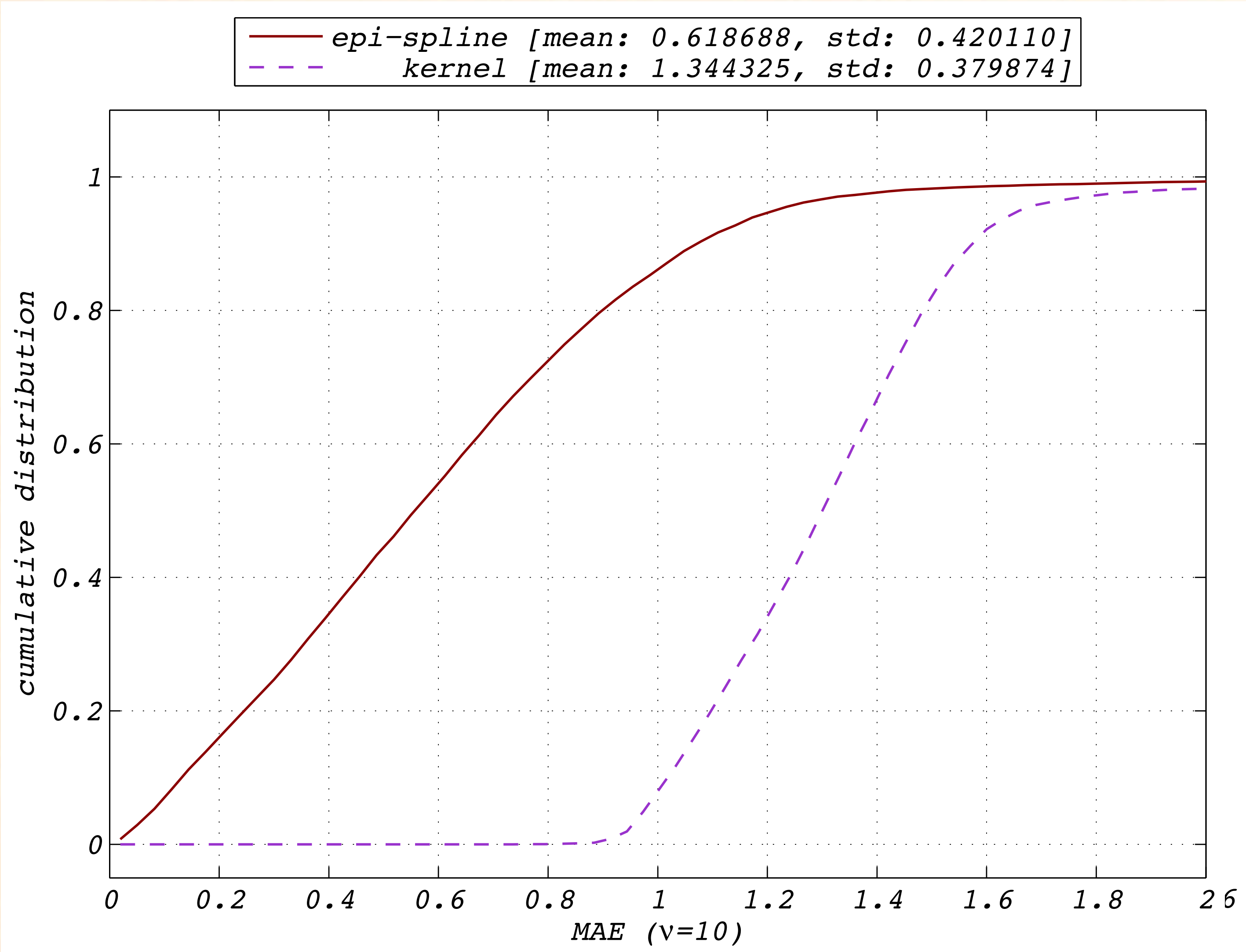


Mean square error
samples: 10
runs: 10,000
kernel estim: 0.293
epi-spline est.: 0.083

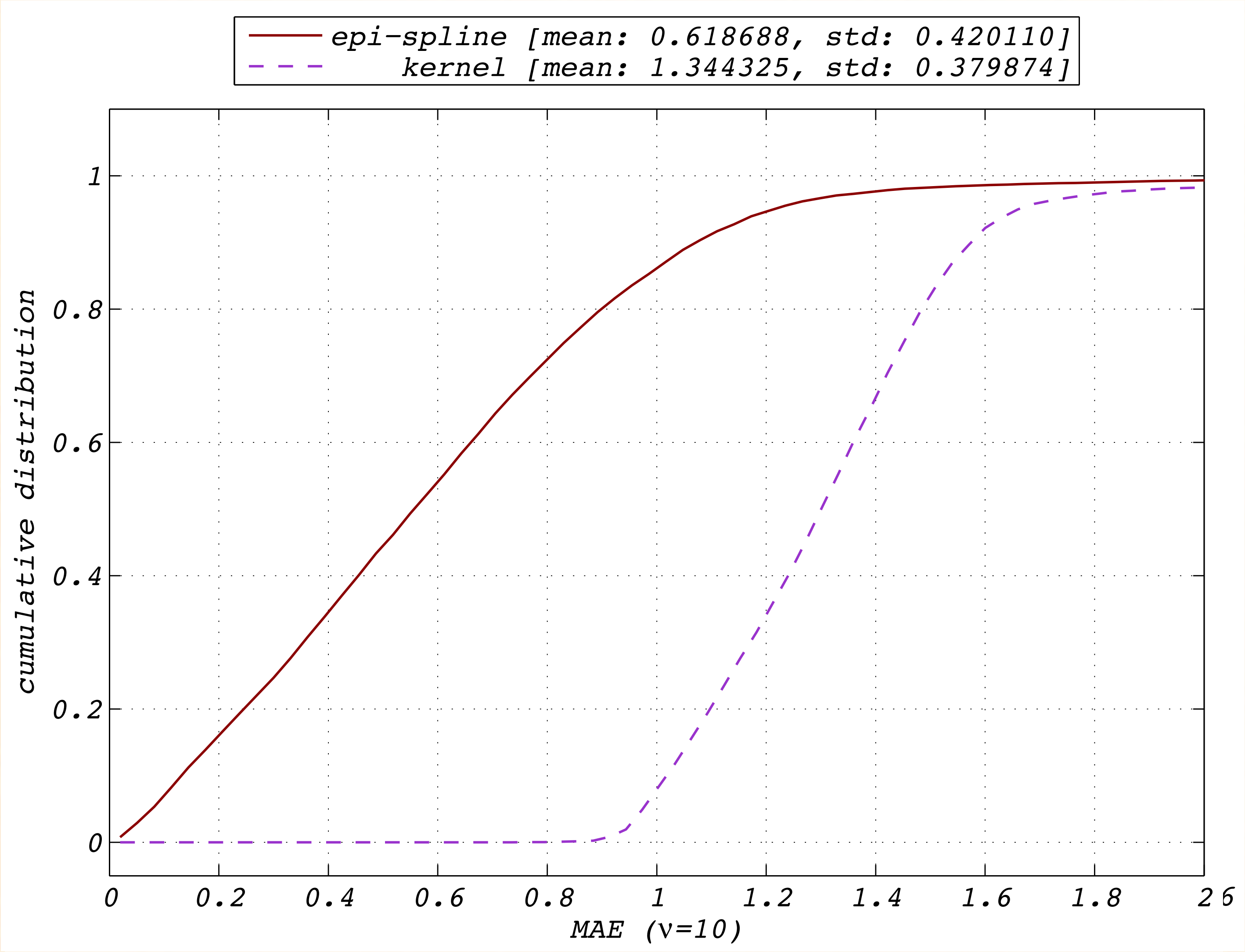
DOES IT PAY OFF? ERROR ANALYSIS



DOES IT PAY OFF? ERROR ANALYSIS

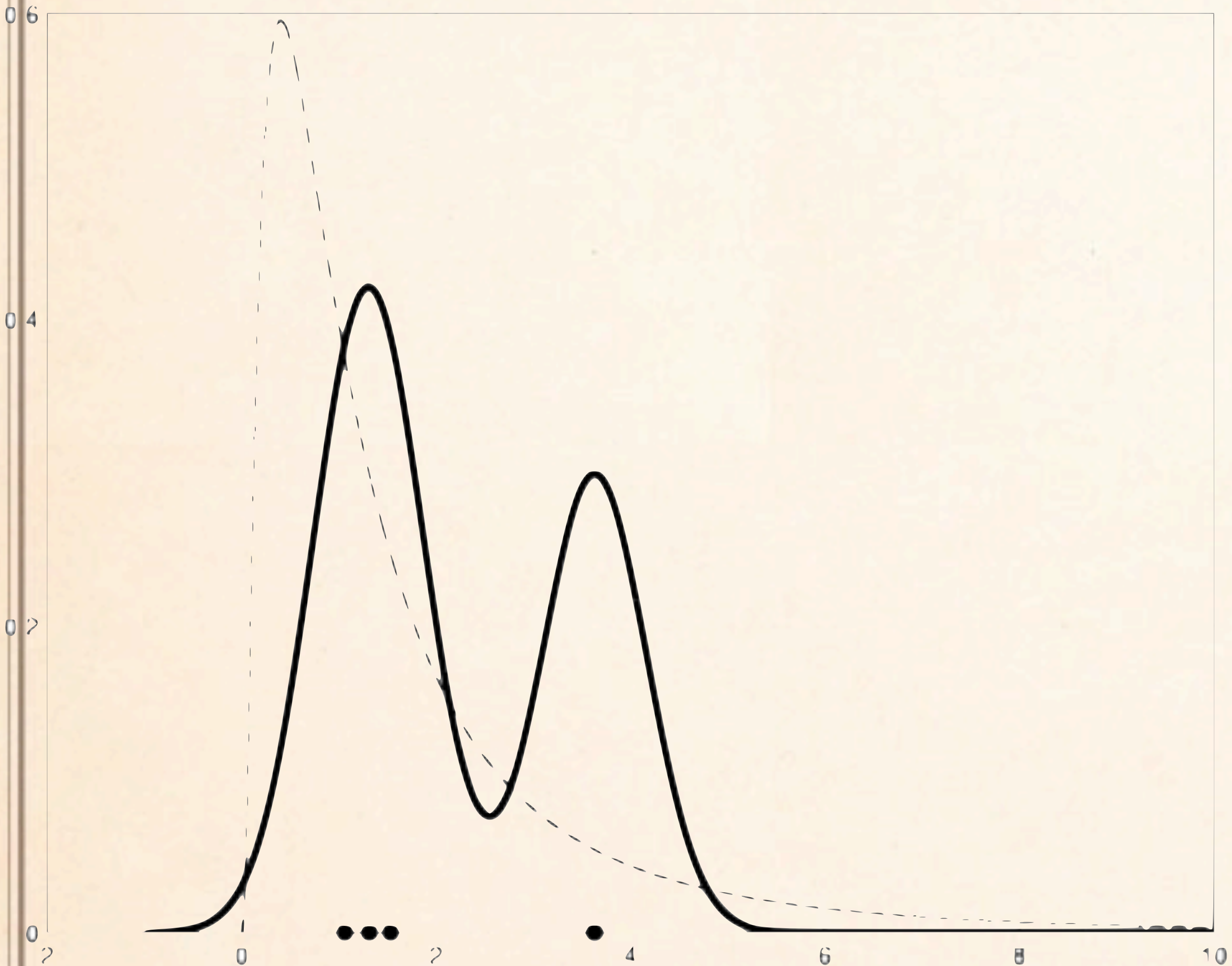


DOES IT PAY OFF? ERROR ANALYSIS



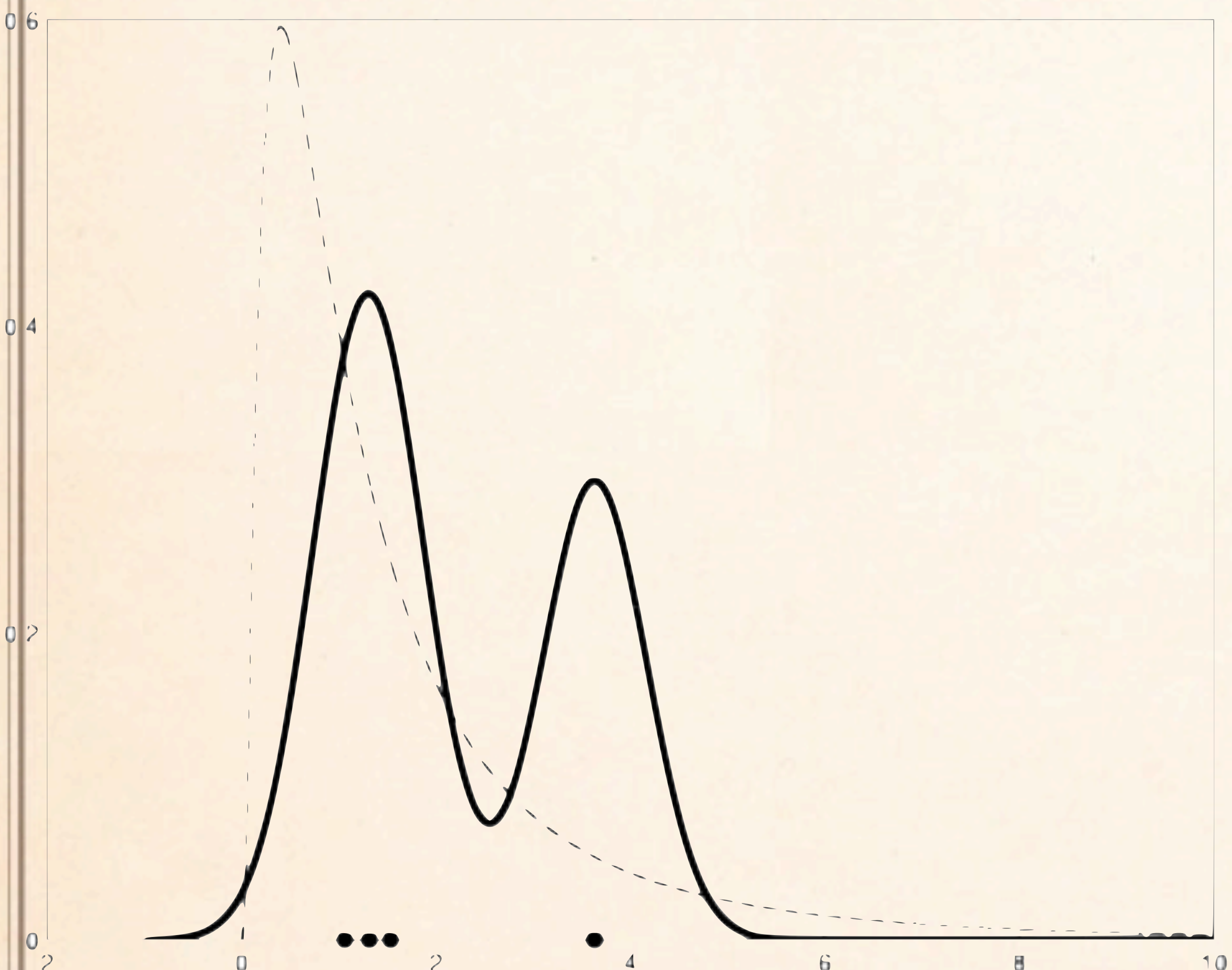
Max. absolute error
samples: 10
runs: 10,000
kernel estim: 1.344
epi-spline est.: 0.619

LOG-GAUSSIAN, N = 4!

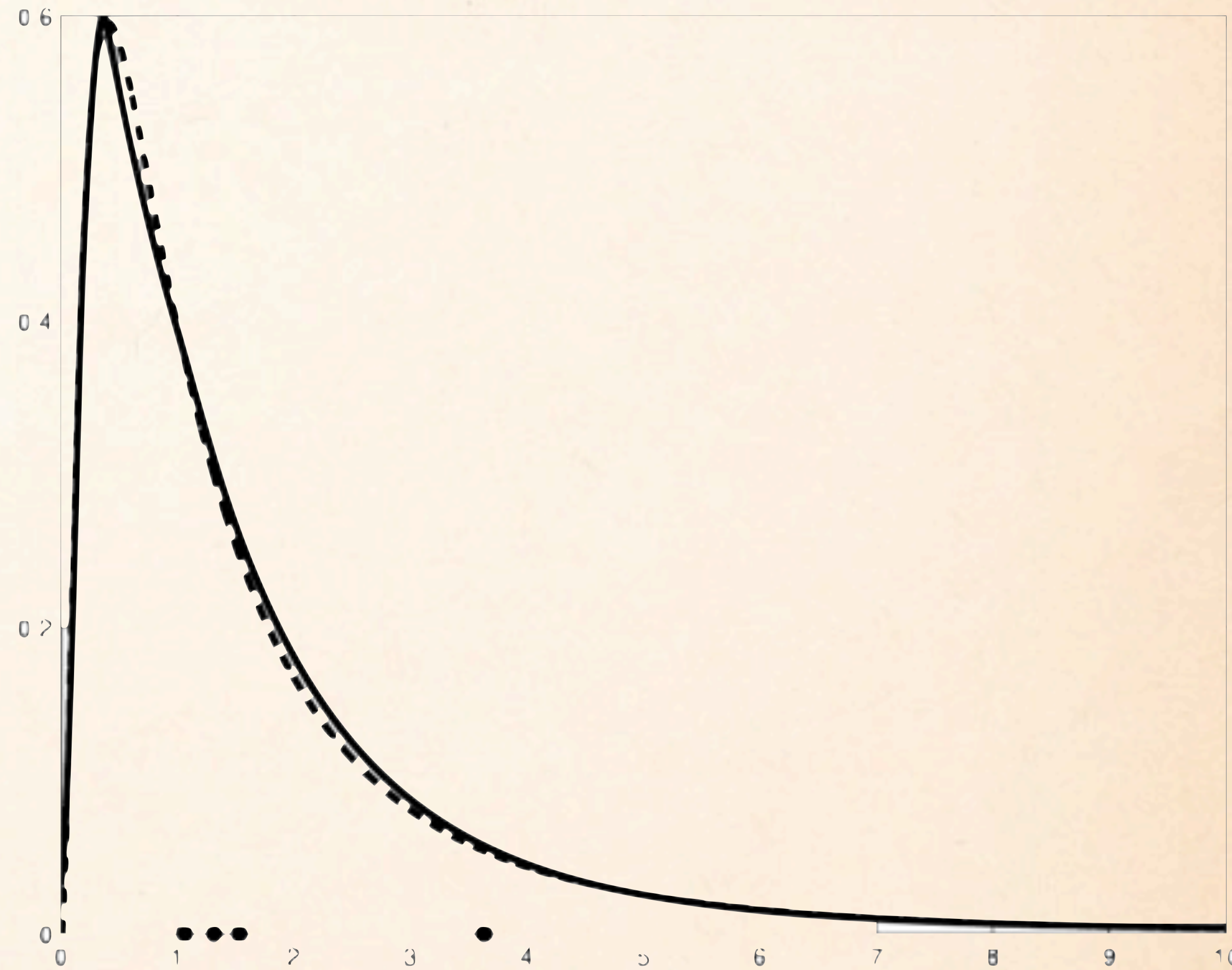


R-Stat (kernel) estimate

LOG-GAUSSIAN, N = 4!



R-Stat (kernel) estimate



support + unimodal

EXTENSIVE EXPERIMENTATION

www.math.ucdavis.edu/~prop01

EPI-SPLINES
An Approximation Tool

EPI-SPLINE: 2nd Order

$c:(a,b] \to \mathbb{R}$ twice differentiable (not C^2)

$c'':(a,b] \to \mathbb{R}$ $2nd$ derivative approzimated by $z:\mathbb{R} \to \mathbb{R}$

split (a,b]: $\{(x_{k-1},x_k],k=1,\dots,N\}$, N relatively large

fix $z(t)=z_k$ (constant) for $t \in (x_{k-1},x_k]$

$2nd$ order epi-spline (1-dim.) +++ mesh = $\max_{k=1,\dots,N} |x_k - x_{k-1}| = m$

$$s_m(x) = s_0 + v_0 x + \int_0^x dr \int_0^r dt\, z(t), \quad z(t) \equiv z_k \text{ on } (x_k, x_{k+1}]$$

$$= s_0 + v_0 x + \sum_{j=1}^{k} a_{kj} z_j \quad \text{when } x \in (x_k, x_{k+1}]$$

$s_m \in C^{1,+pl}$, linear w.r.t. $s_0, v_0, z_1, \dots, z_N$ (finite # parameters)

as mesh $m \searrow 0$, $\max \|s_m - c\|^2 \to 0$ and s_m epi-converges to c

EPI-SPLINES

originally (Wets, Bianchi & Yang, 2002):

derive financial curves,

later, also stochastic volatility

Epi-splines of k*th* order:

piece-constant k*th* derivative

still linear w.r.t. its parameters

and epi-convergence to *c*

EXPONENTIAL EPI-SPLINE
1-dimensional, 2nd order

$$h(x) = e^{-s(x)}$$

$$s(x) = s_0 + v_0 x + \int_0^x dr \int_0^r dt\, z(t), \quad z(t) \equiv z_k \text{ on } (x_k, x_{k+1}]$$

$$= s_0 + v_0 x + \sum_{j=1}^{k} a_{kj} z_j \quad \text{when } x \in (x_k, x_{k+1}]$$

$$\max E^\nu \left[\ln h(\xi) \right] = \frac{1}{\nu} \sum_{l=1}^{\nu} \ln h(\xi^l) = \min \frac{1}{\nu} \sum_{l=1}^{\nu} s(\xi^l)$$

$$\text{such that } \int e^{-s(\xi)}\, d\xi \le 1. \qquad (h \ge 0)$$

$$z_k \in \left[-\kappa_l, \kappa_u \right] \quad \text{'constrained' curvature}$$

$$\text{unimodal: } \kappa_l = 0 \Rightarrow s(\cdot) \text{ convex}$$

HIGHER-DIMENSIONAL EPI-SPLINES
2-DIMENSIONAL - FIRST VERSION

$$s(x,y) = z_0 + v_1 x + v_2 y + \int_0^x d\tau \int_0^\tau d\theta\, a(\theta)$$

$$+ \int_0^y d\tau \int_0^\tau d\theta\, b(\theta) + \int_0^x d\tau \int_0^y d\theta\, c(\tau,\theta)$$

on $(x_{k-1}, x_k] : a(x) = a_k$,

on $(y_{k-1}, y_k] : b(y) = b_k$,

on $(x_{k-1}, x_k] \times (y_{l-1}, y_l) : c(x,y) = c_{kl}$

requires bouundary continuity properties

estimation : a_k and $b_k \Rightarrow$ marginal distributions

c_{kl} correlation coefficients (locally)

HIGHER-DIMENSIONAL EPI-SPLINES
2-DIMENSIONAL - FIRST VERSION

$$\nabla^2 s(x) = \begin{pmatrix} a_{k-1,l-1} & a_{k-1,l} \\ a_{k,l-1} & a_{k,l} \end{pmatrix} \text{ Hessian}$$

on open rectangle $(x_{k-1}, x_k) \times (y_{l-1}, y_l)$

unimodal $h(x) = e^{-s(x)} \Rightarrow s$ convex (globally)

$\nabla^2 s(x)$ positive semidefinite, symmetric

(3 parameters)

rectangle boundary values: (at a mesh)

via monotonicity of $\nabla s(x)$

all conditions included in the optimization problem

EXAMPLE: NORMAL DENSITY

mean = (0,0) ... data samples correlated

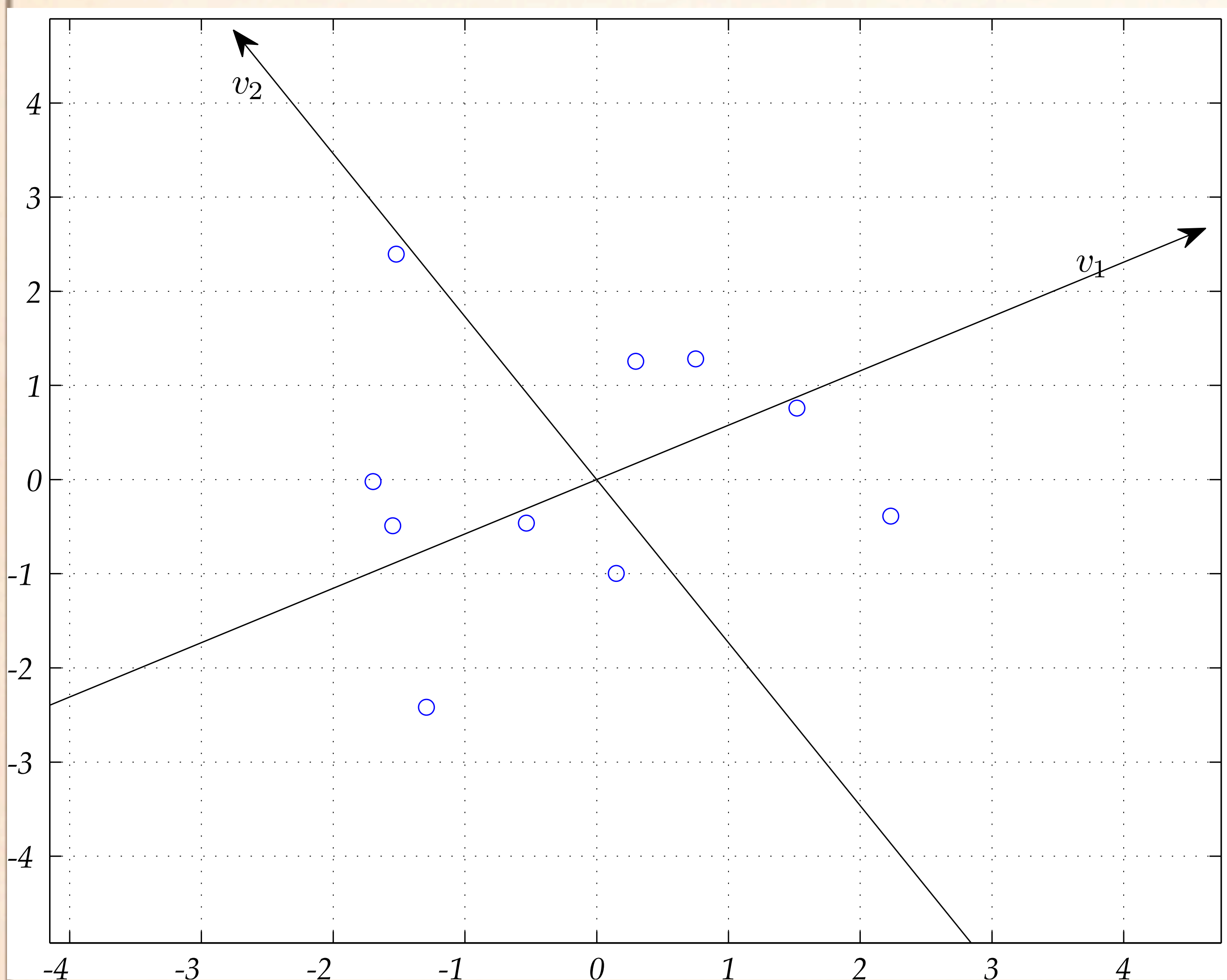
covariance: MDM^T, $D=$ diag(4,1), $M=\begin{pmatrix} \cos(\pi/6) & \cos(2\pi/3) \\ \sin(\pi/6) & \sin(2\pi/3) \end{pmatrix}$

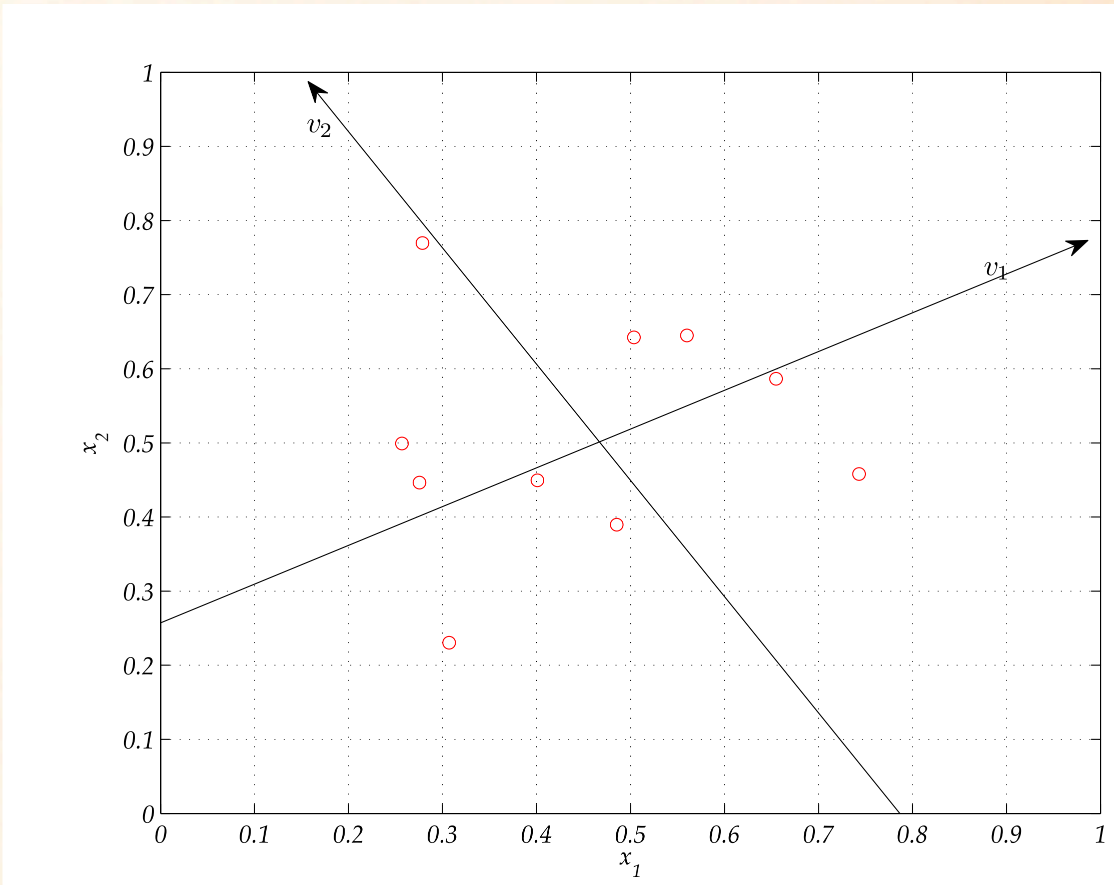
samples: $\nu = 10$, "soft" information: h unimodal

Results:

$$\left\| h^{true} - h^{est} \right\|_2^2 = 0.028, \quad \left\| h^{true} - h^{est} \right\|_\infty = 0.006$$

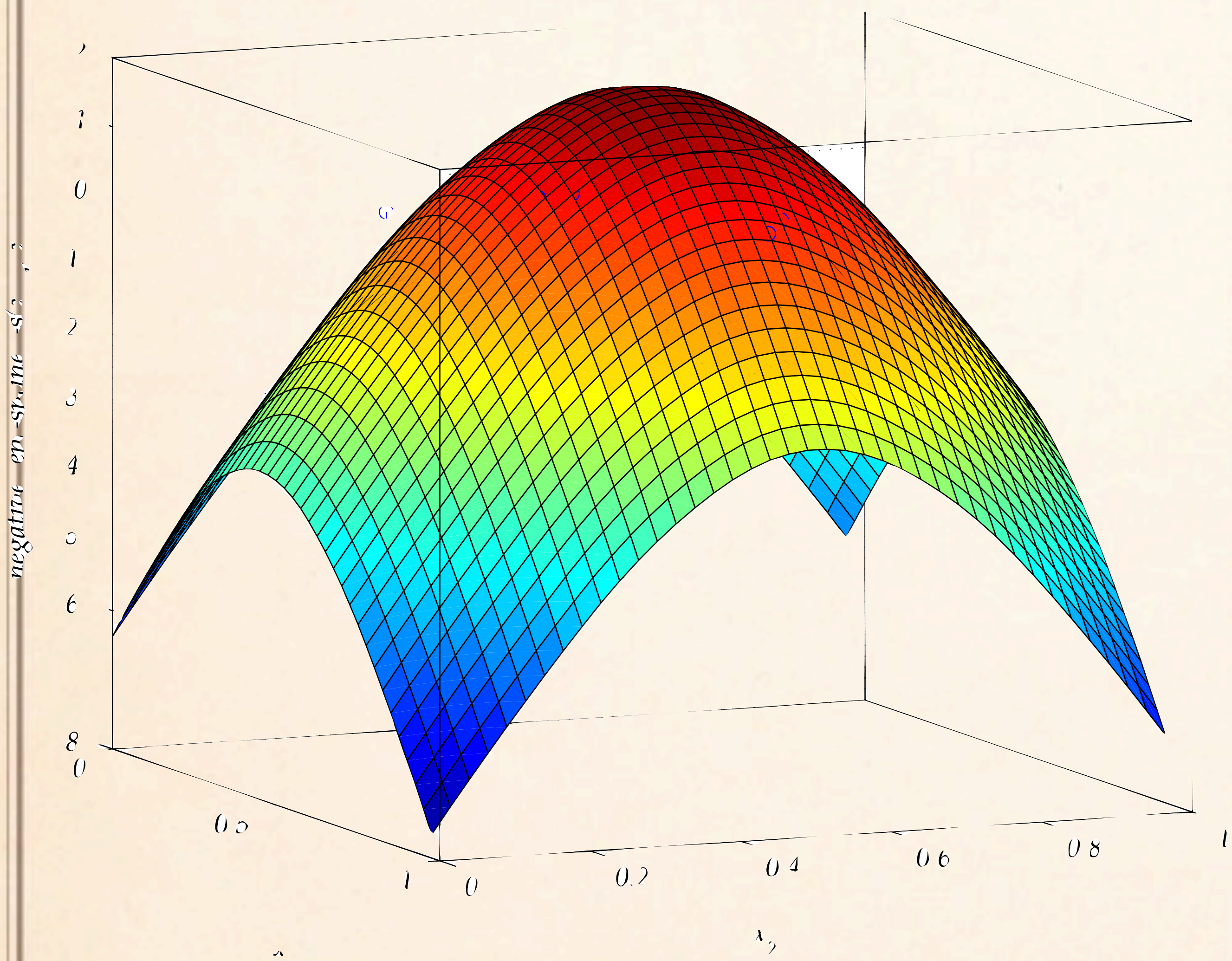
SAMPLED DATA

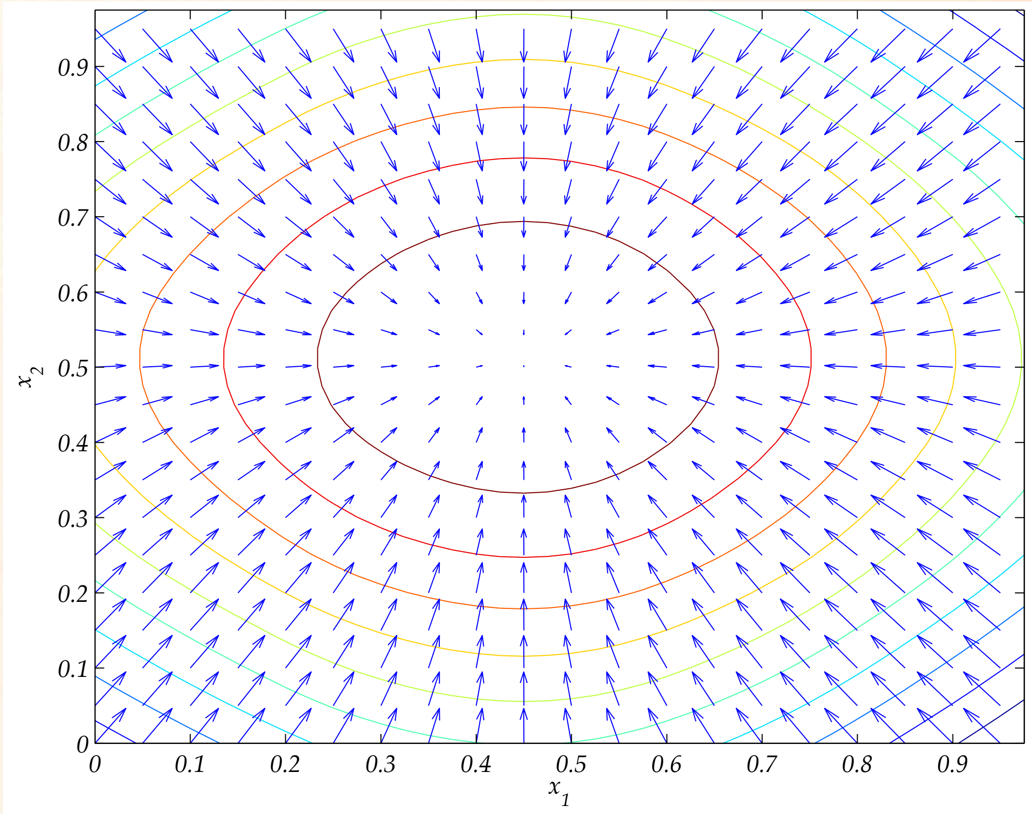




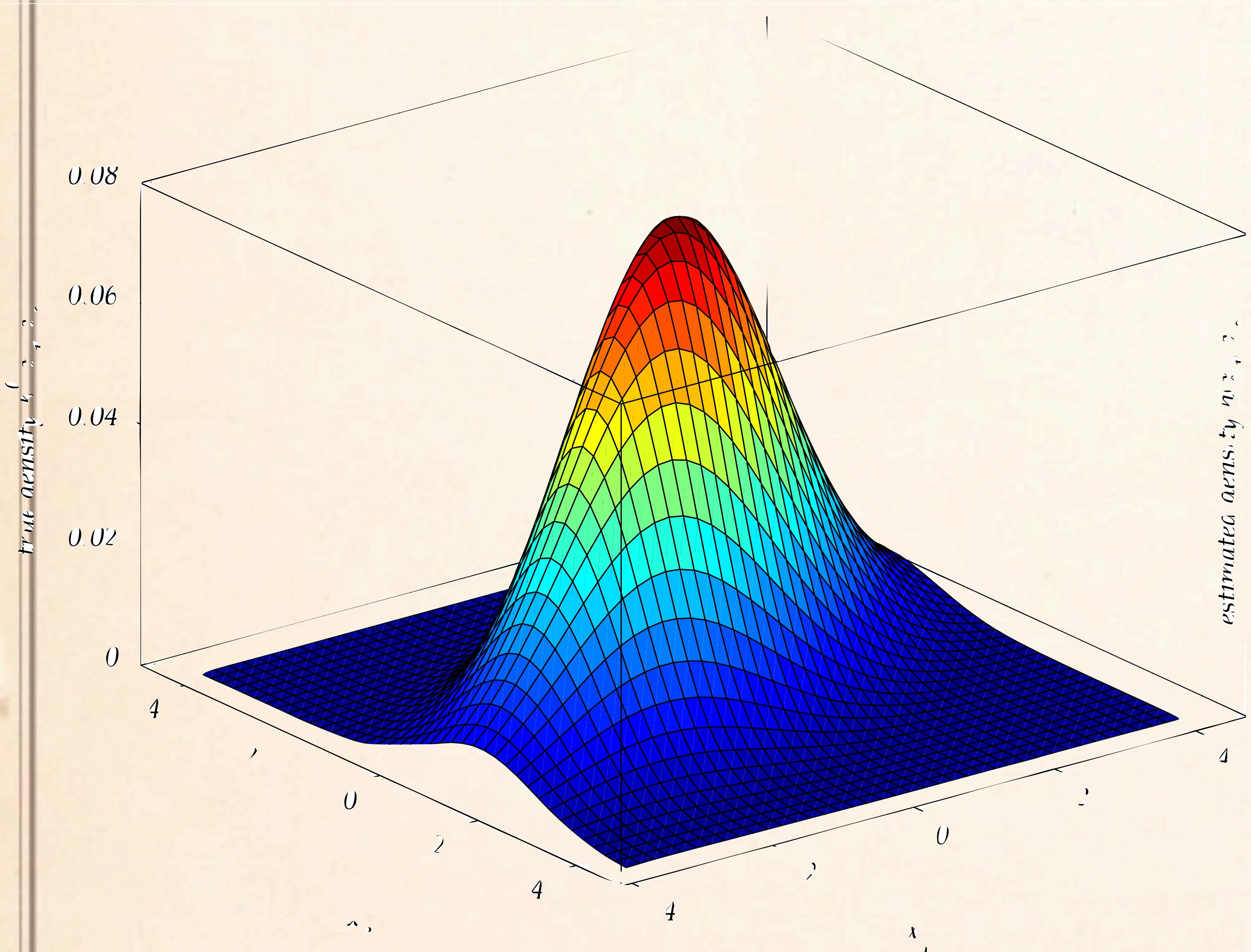
normalized

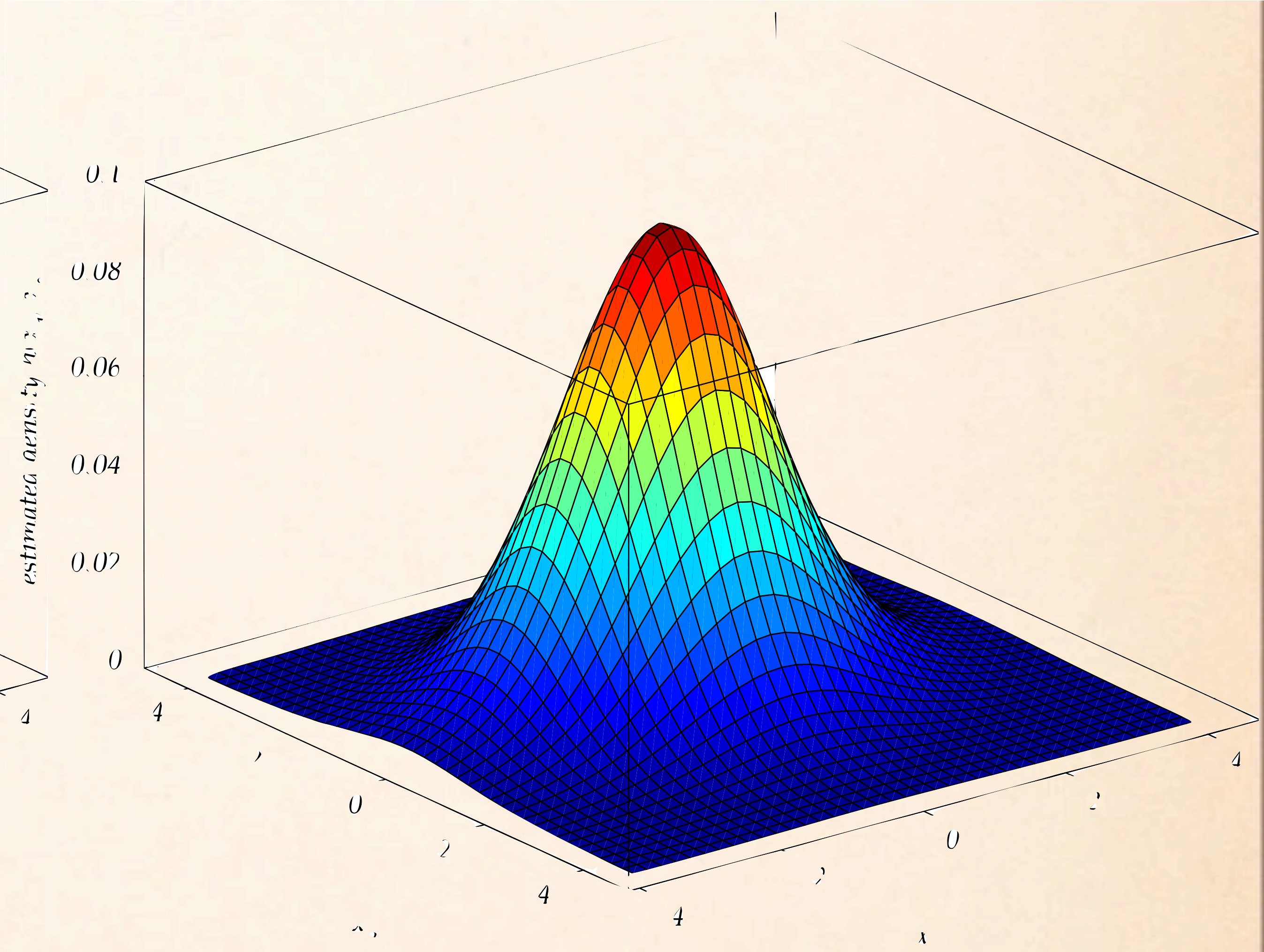
EPI-SPINE & VECTOR FIELD



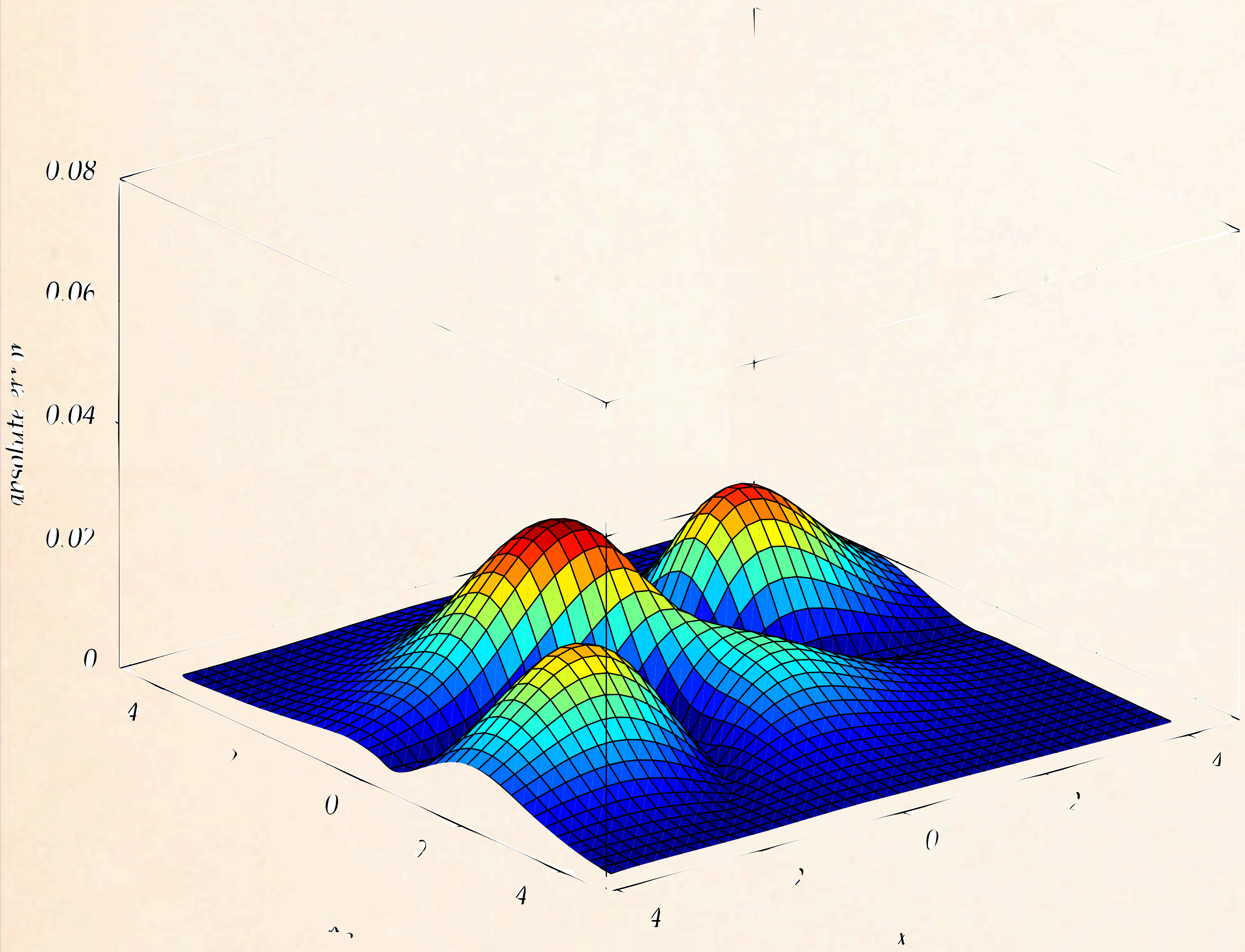


TRUE & ESTIMATED DENSITY

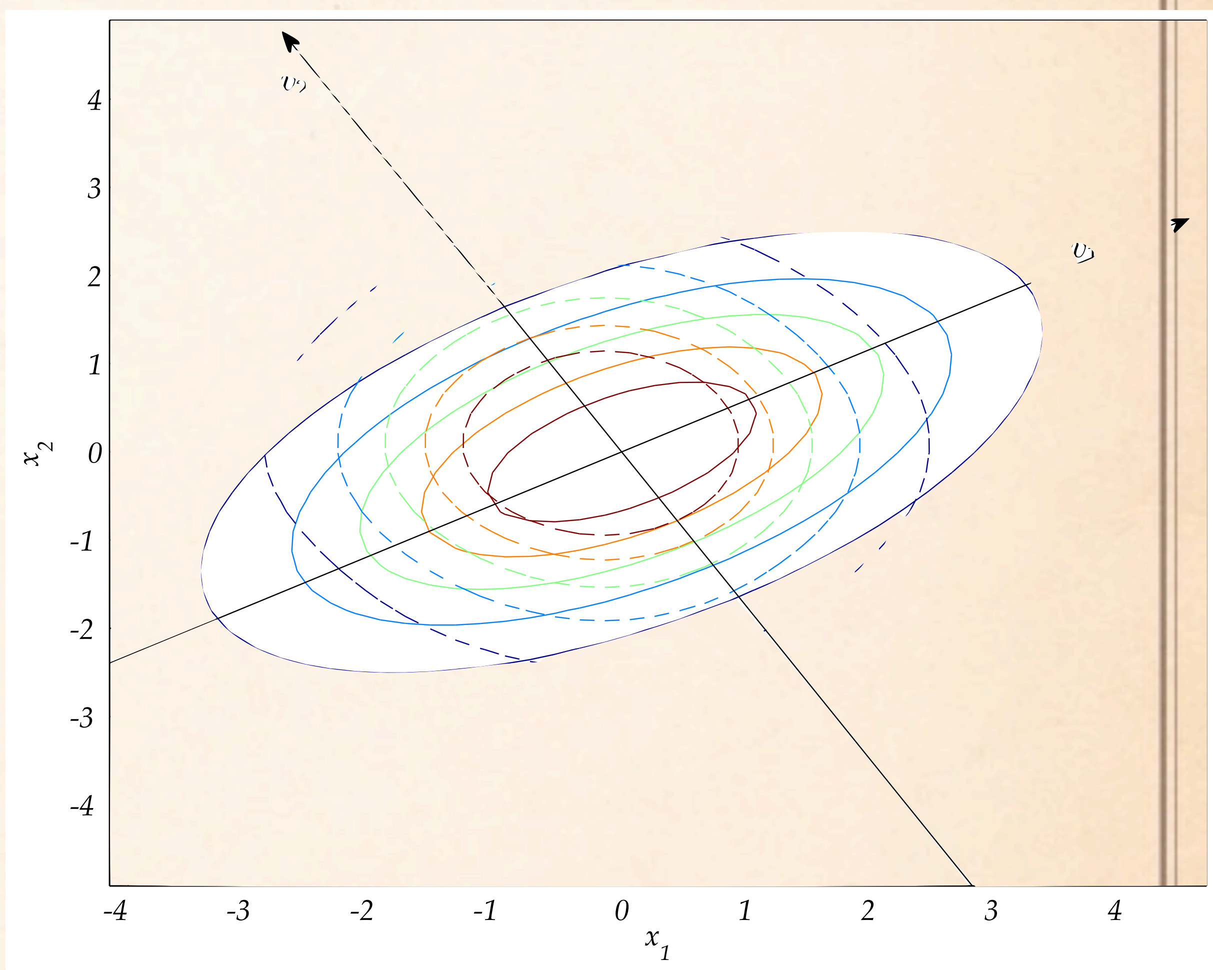




MEASUREMENT ERRORS

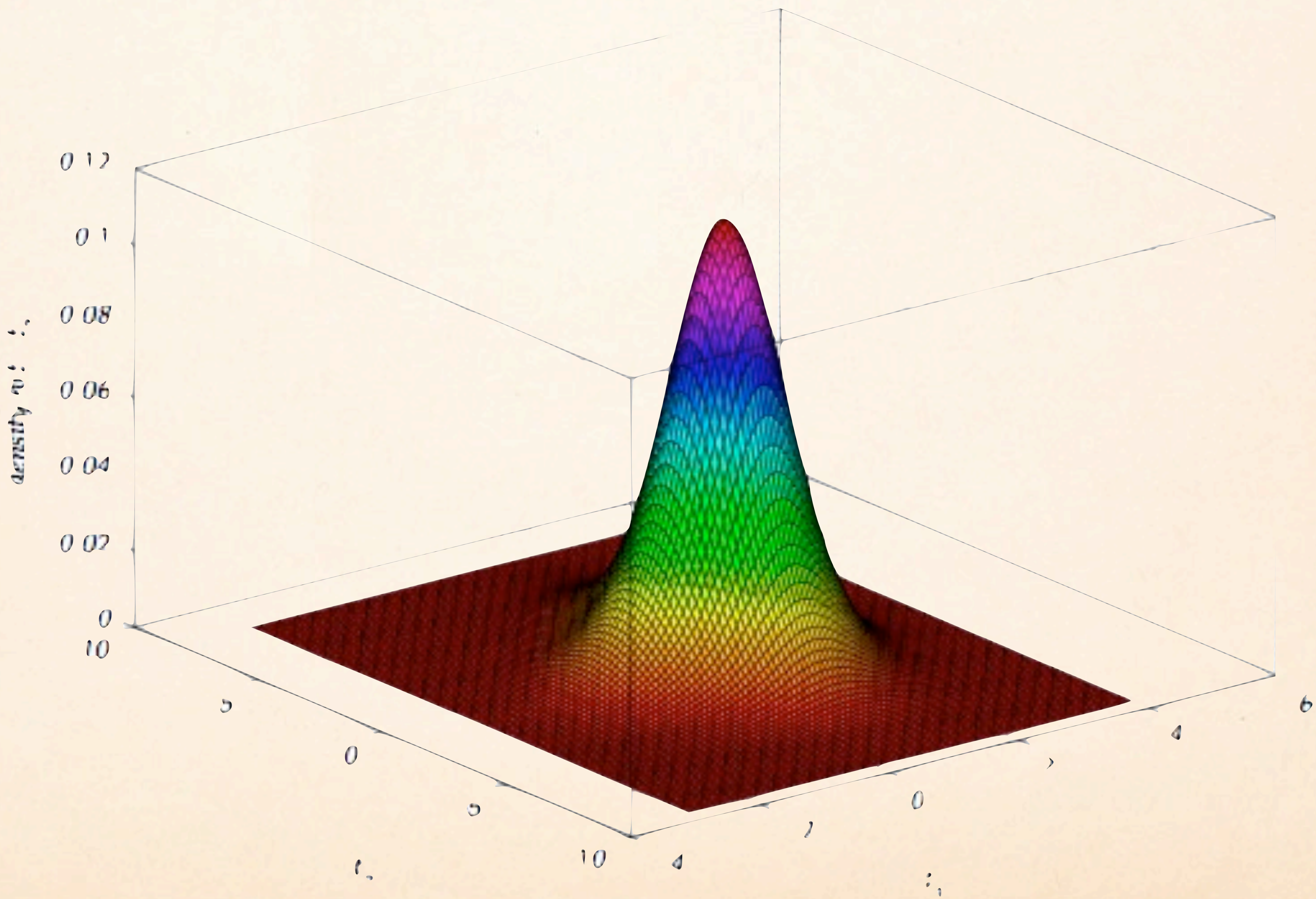


Absolute Error

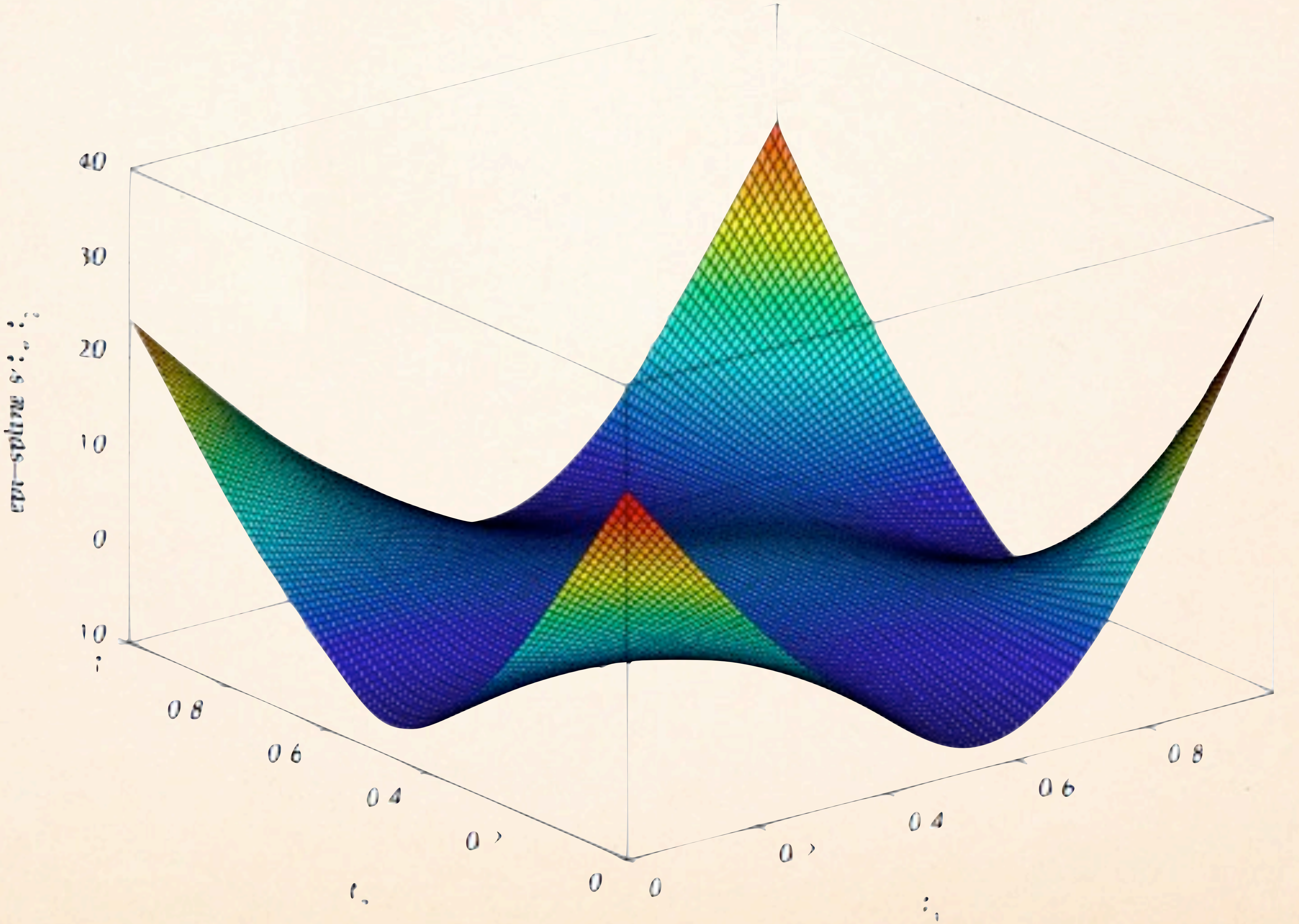


Level curves: true & estimate

NORMAL, 20 SAMPLES DIAGONAL COVARIANCE

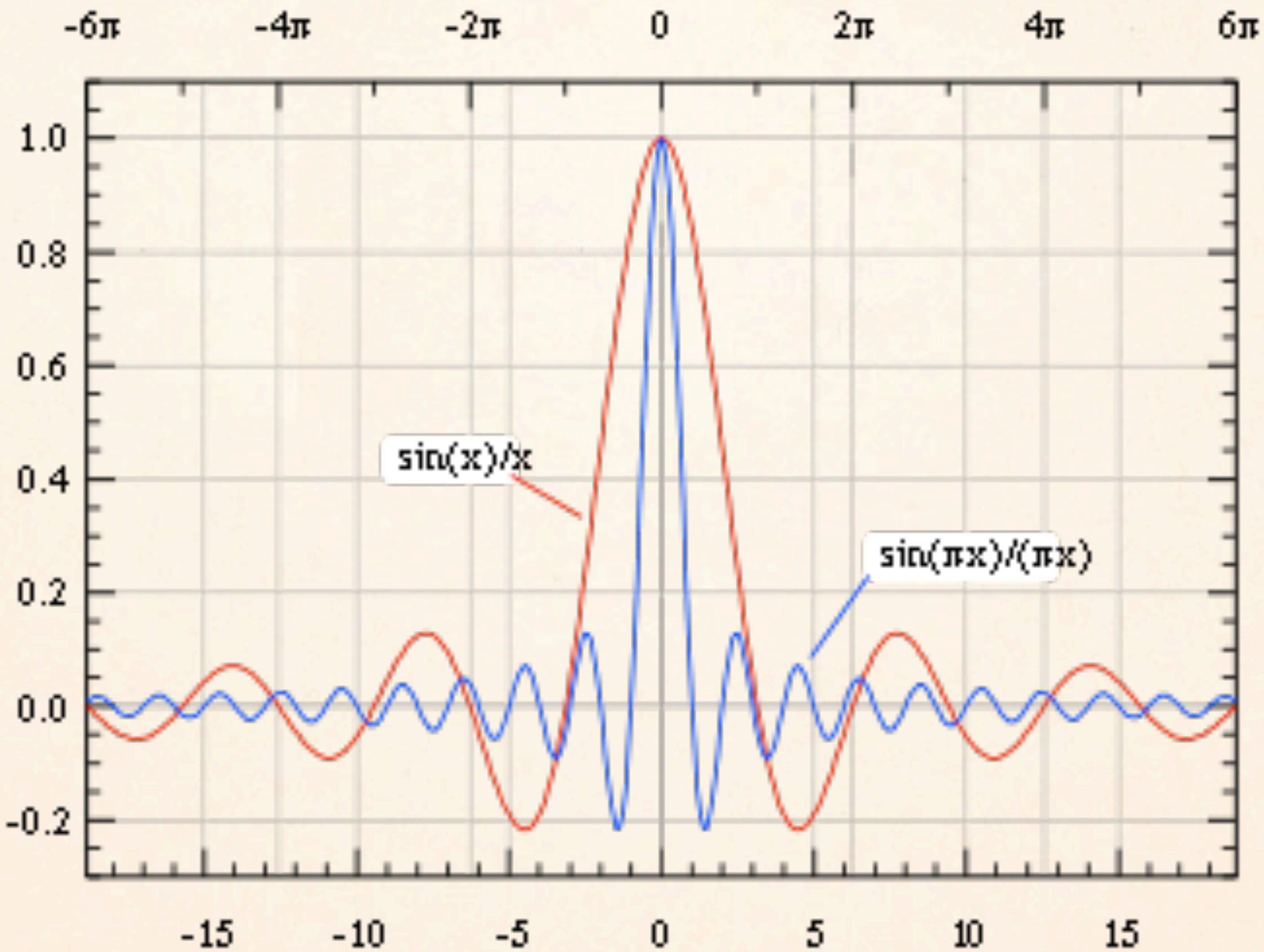


... THE -(EPI-SPLINE FUNCTION)



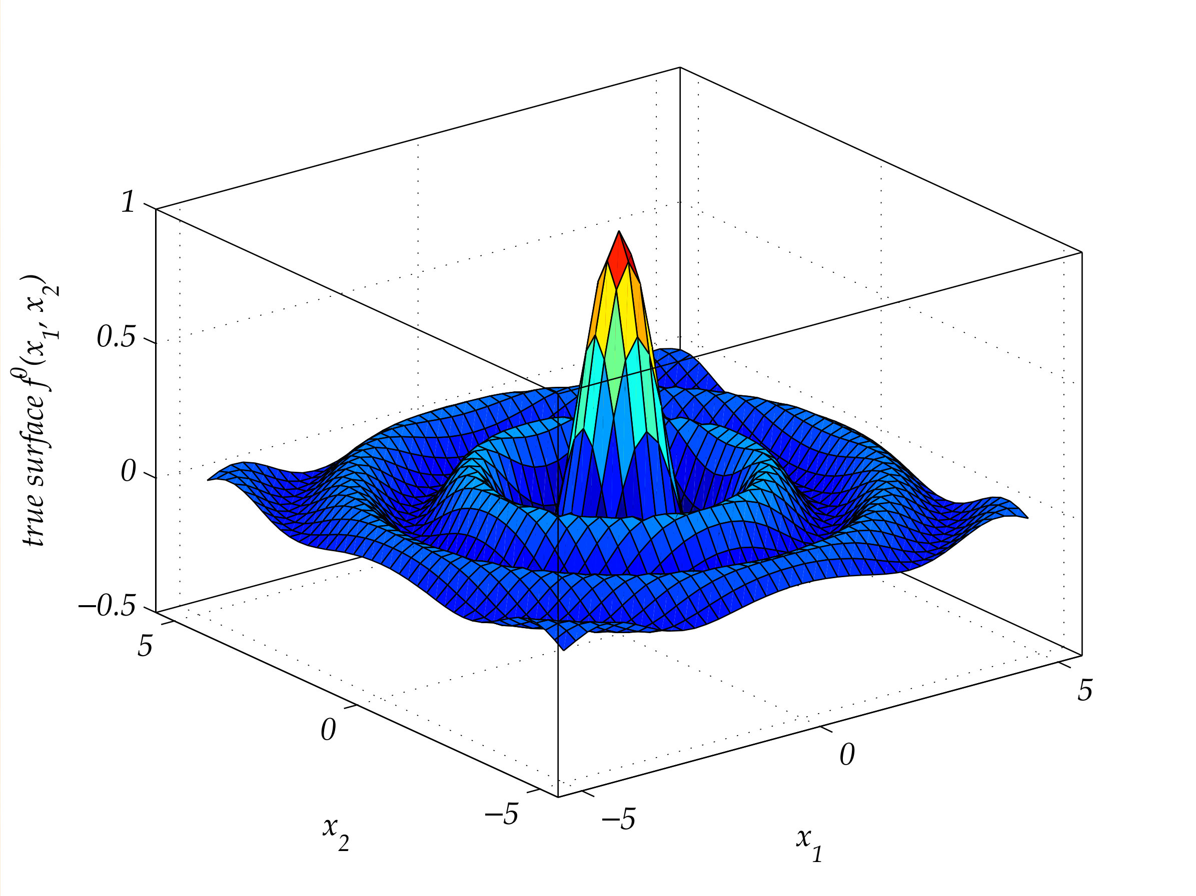
SPLINES & EPI-SPLINES

DIRICHLET (SINC-)FUNCTION

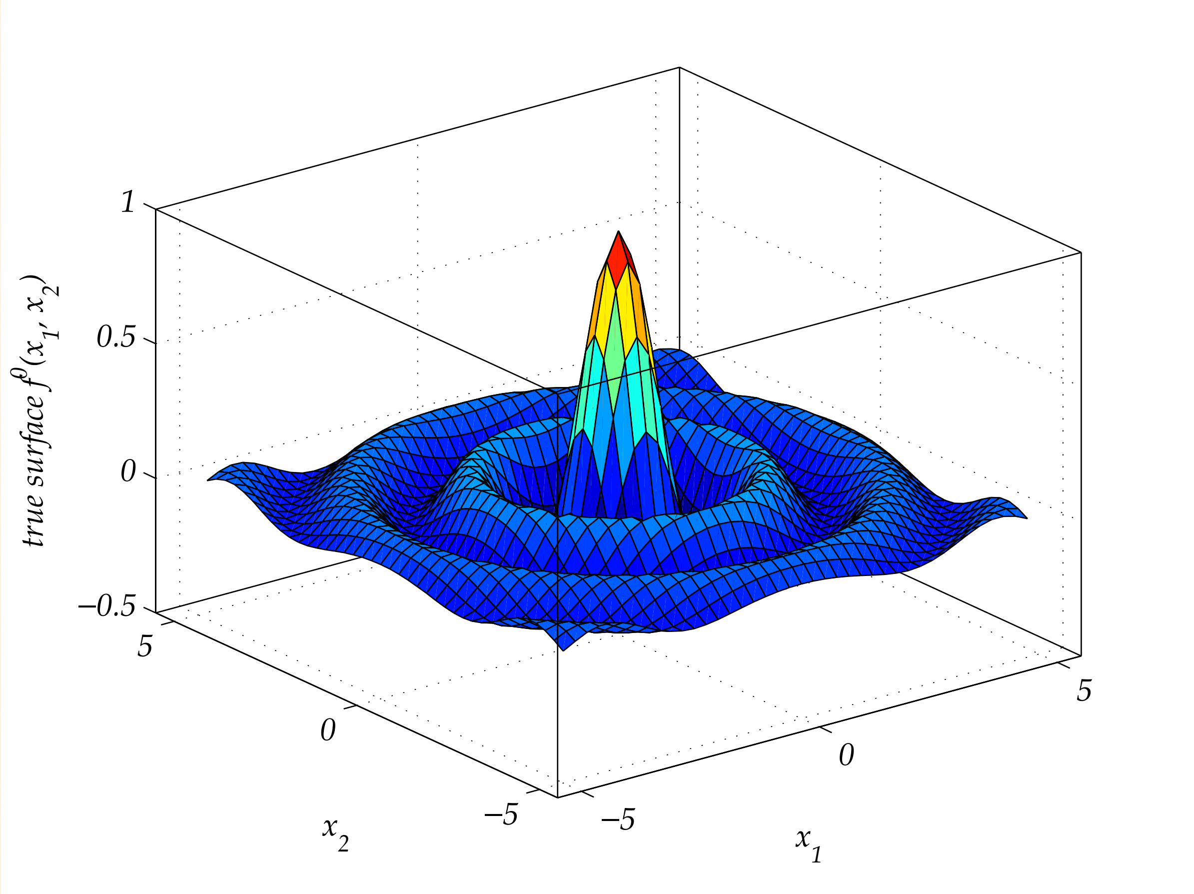


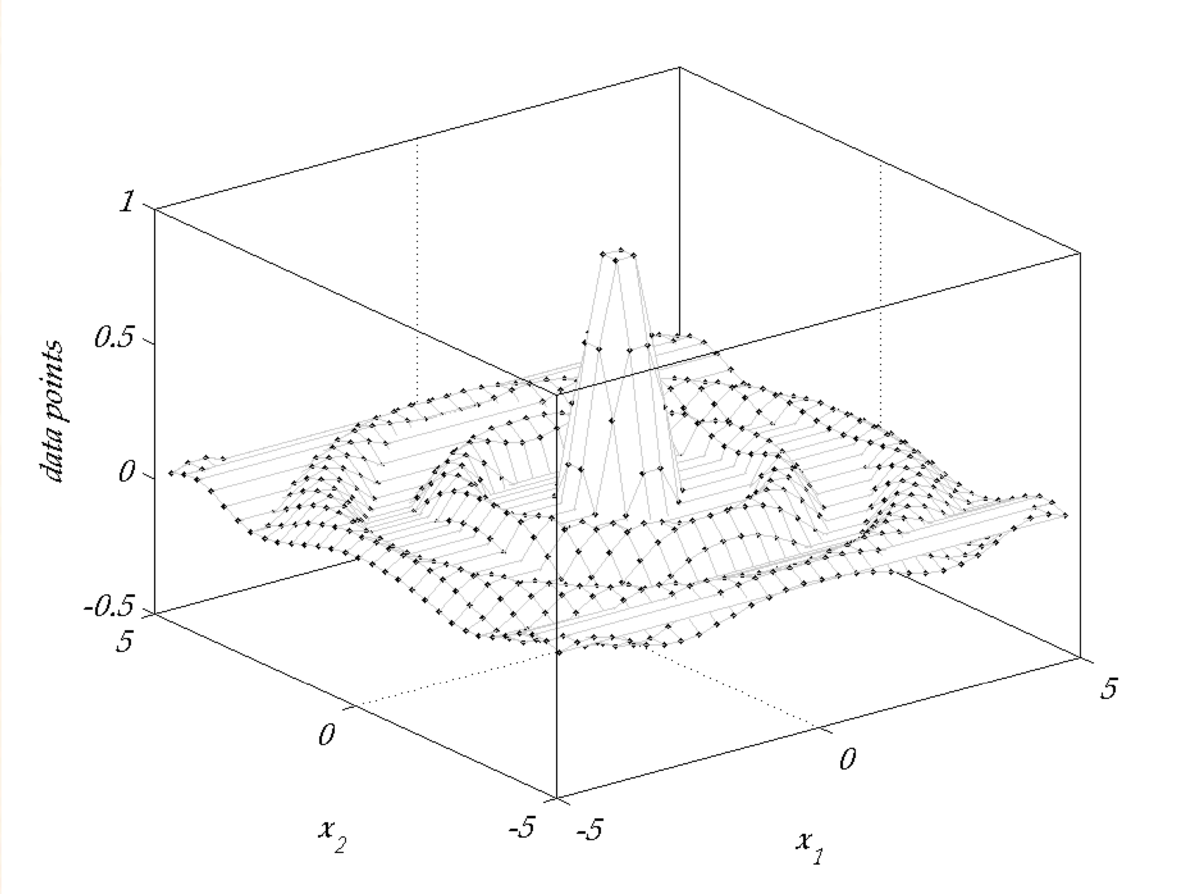
$$f(x) = \sin(\pi x / 2) / \pi x \text{ for } x \neq 0, \quad = 1 \text{ for } x = 0$$

SPLINES & EPI-SPLINES

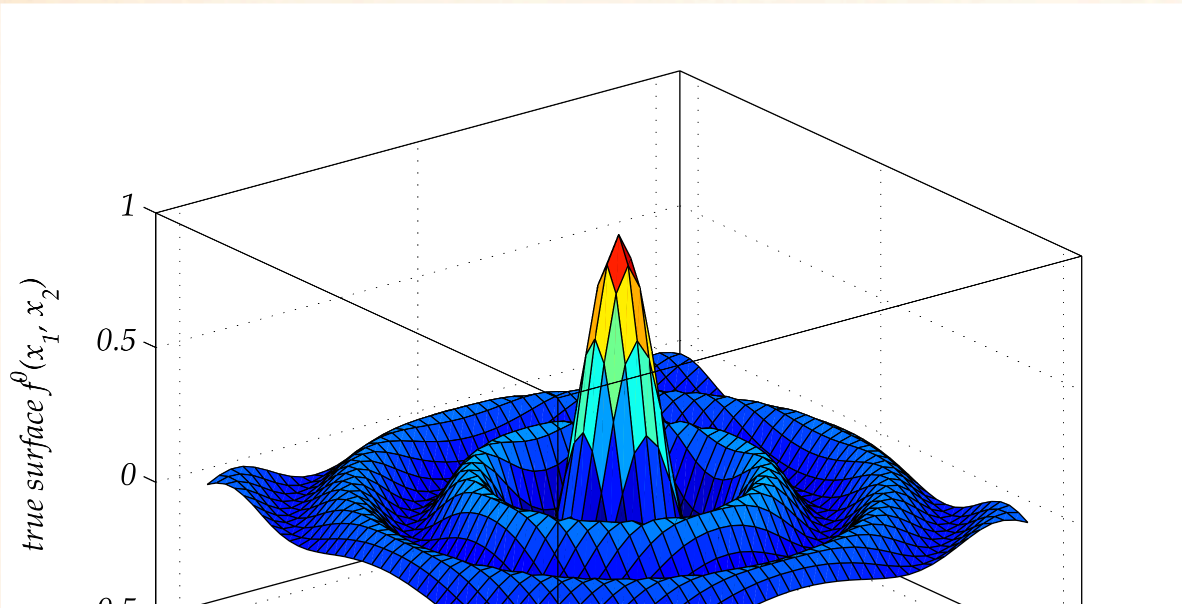


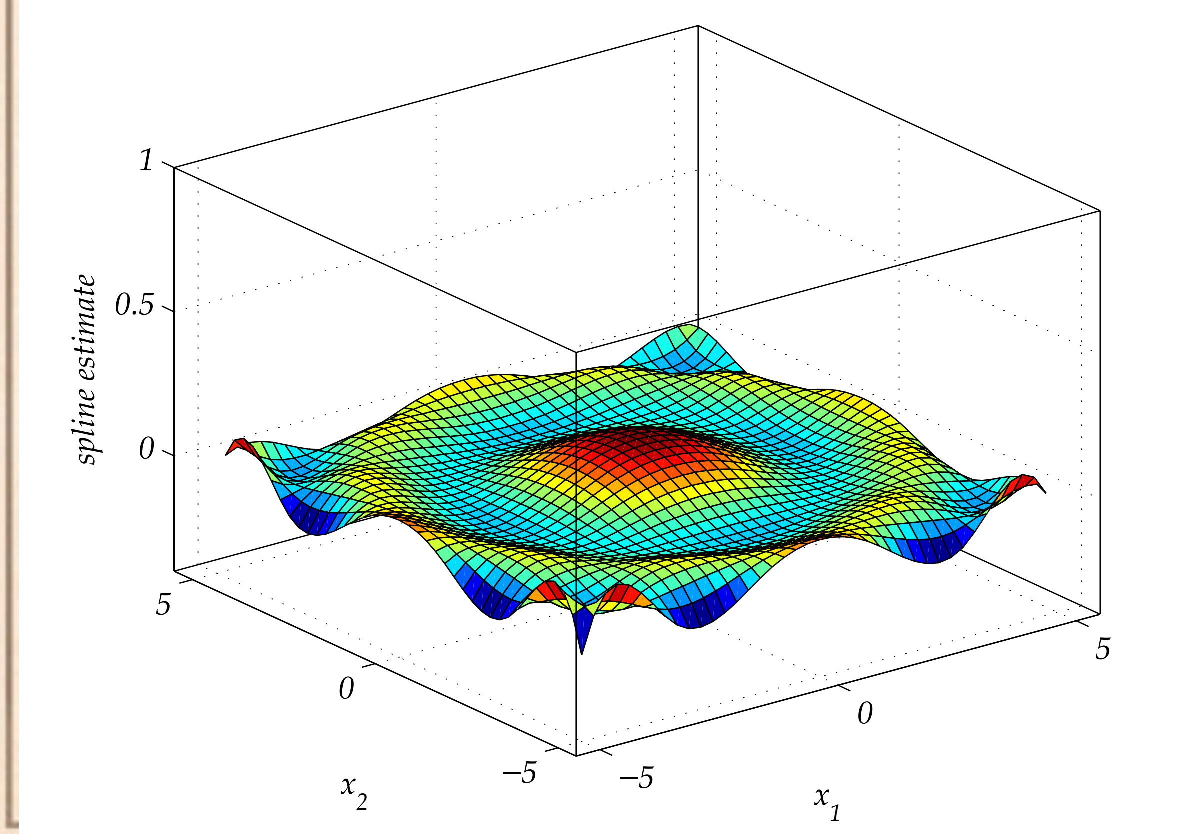
SPLINES & EPI-SPLINES

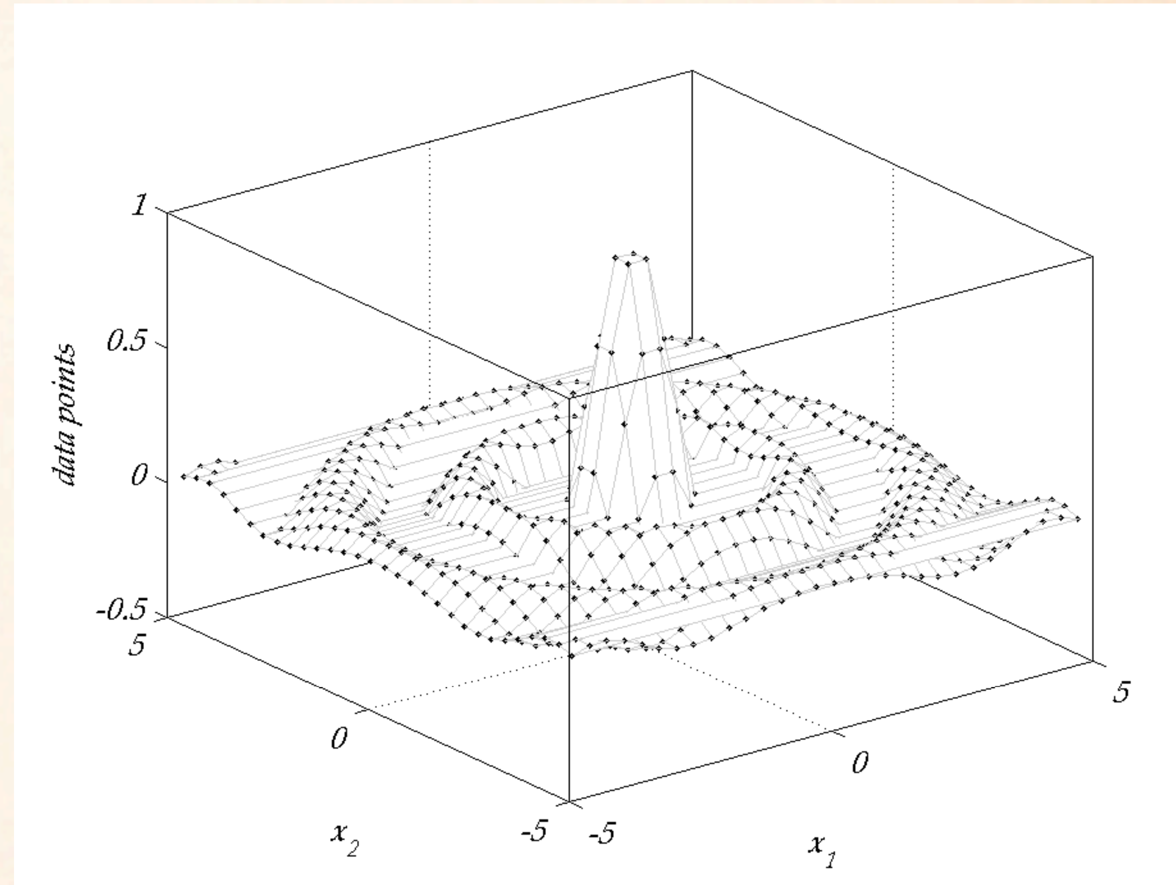




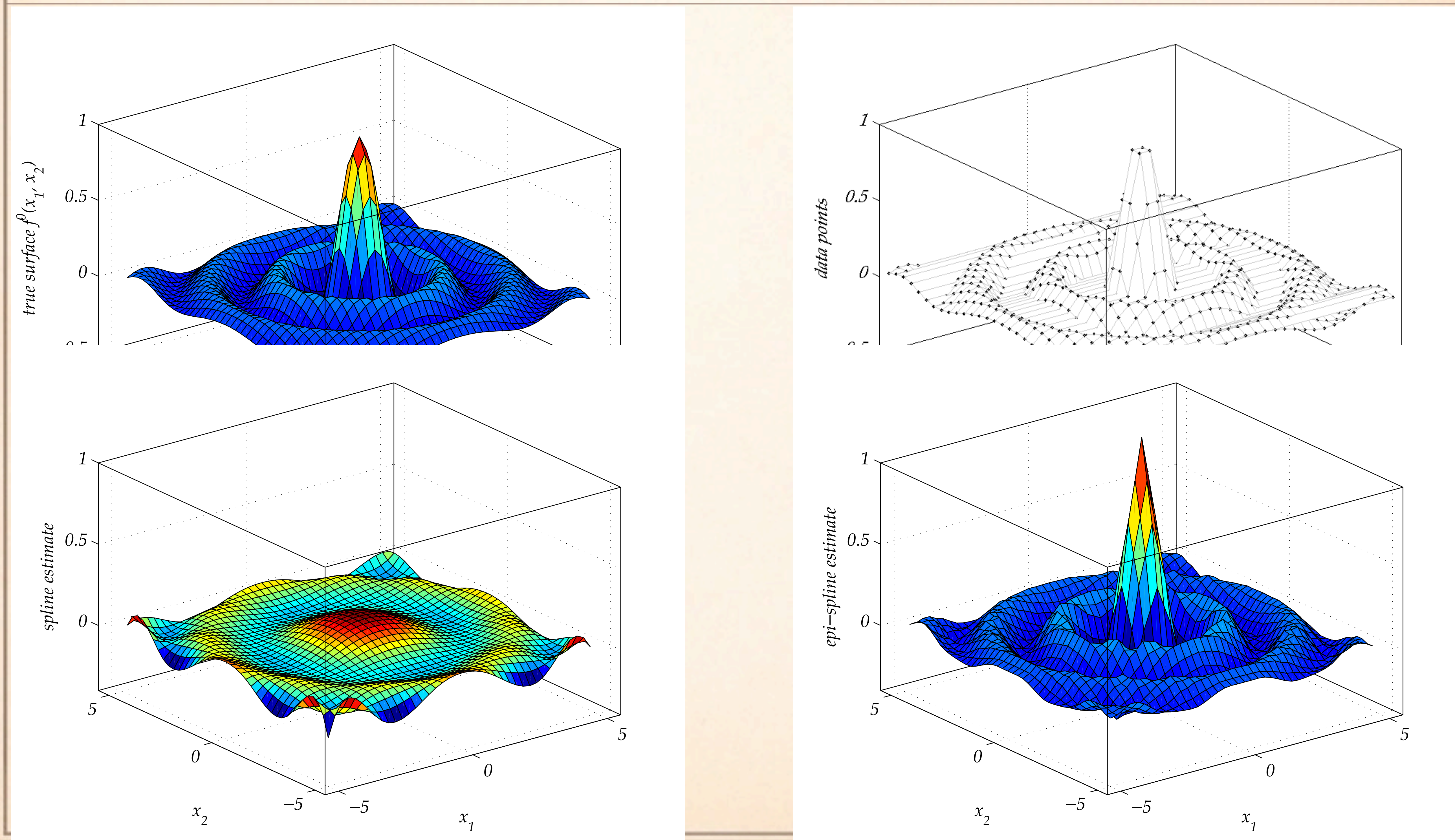
SPLINES & EPI-SPLINES



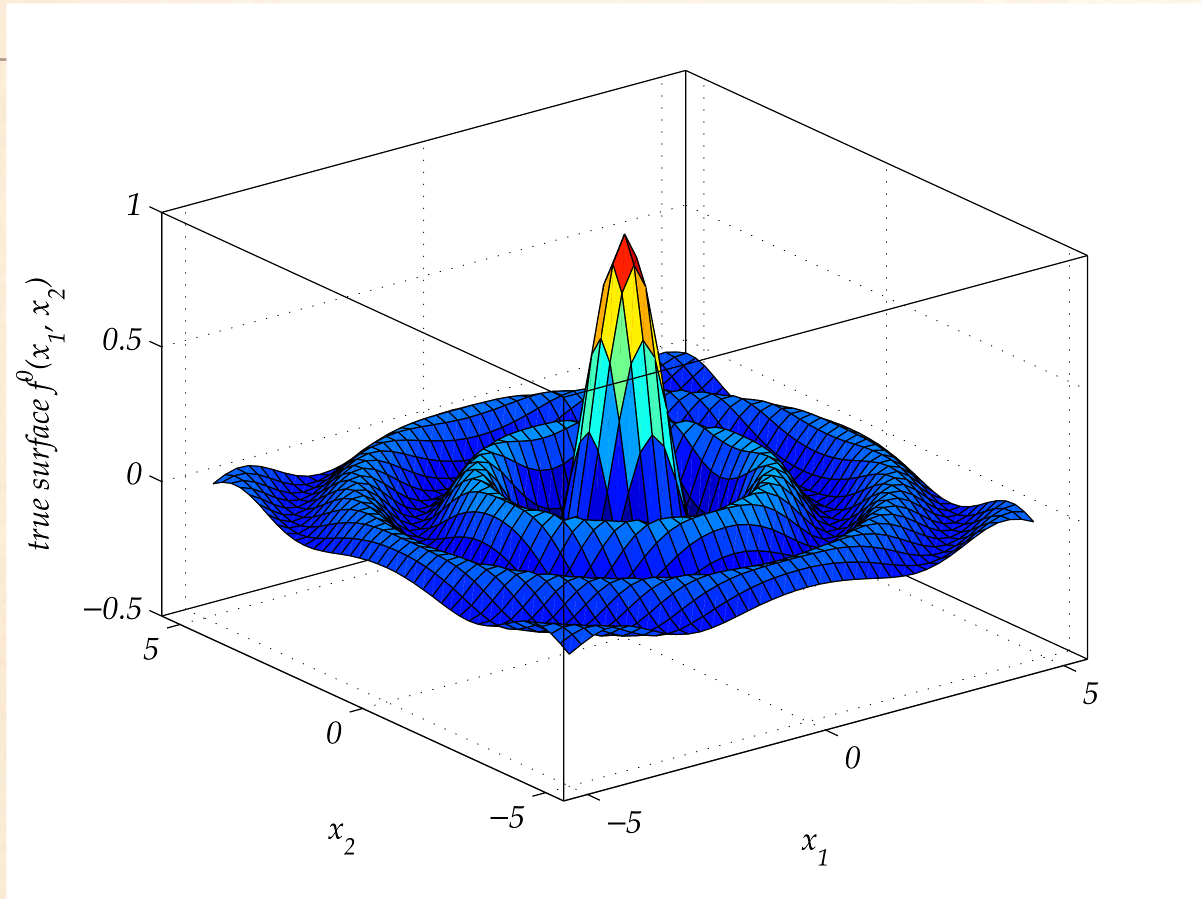




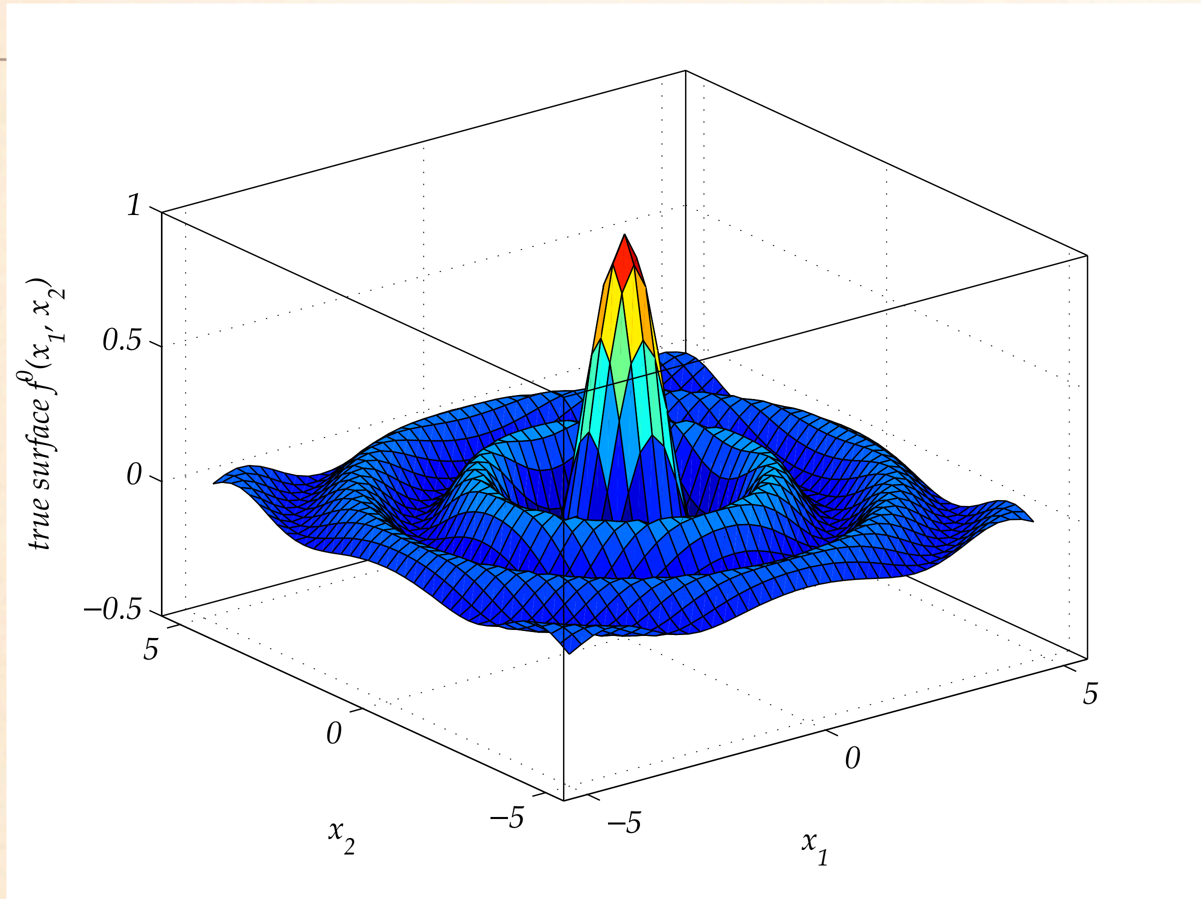
SPLINES & EPI-SPLINES

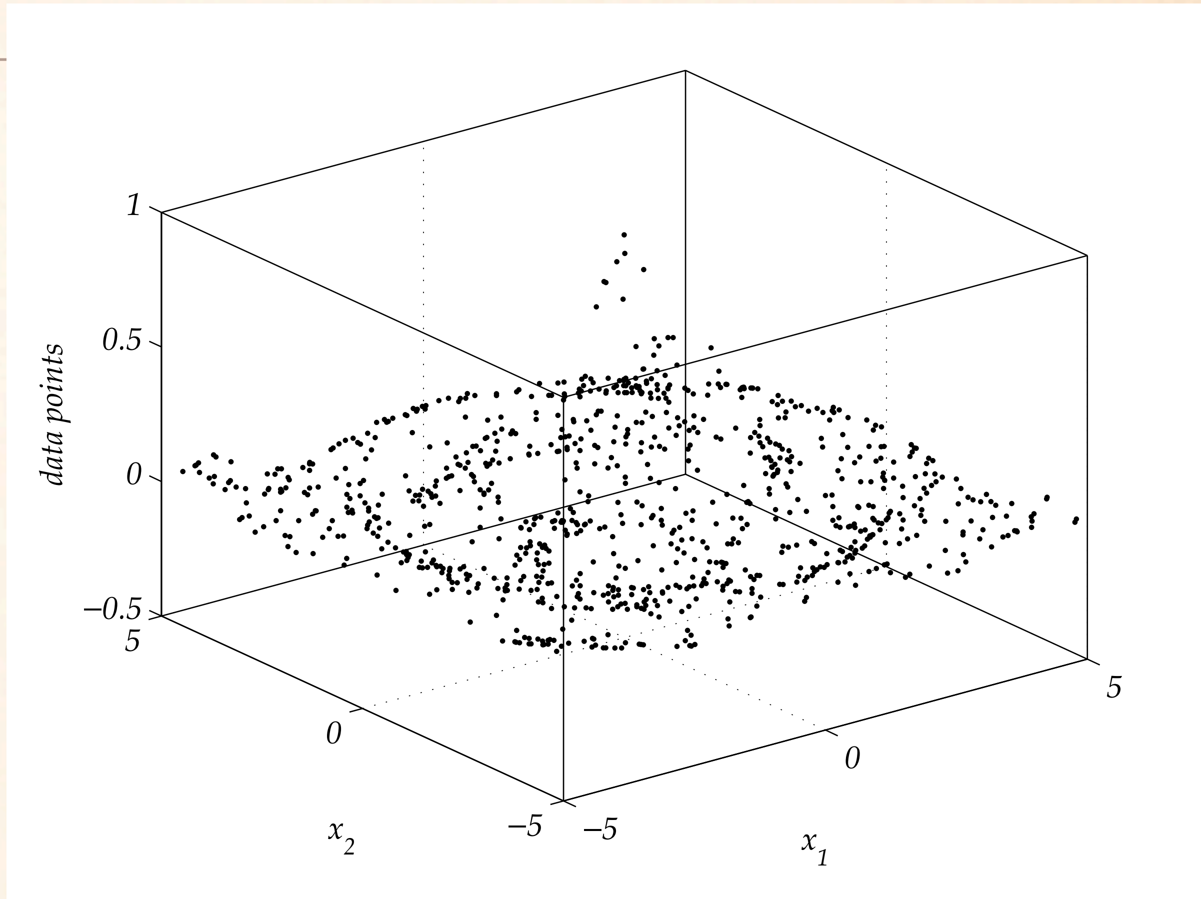


SPLINES & EPI-SPLINES 2

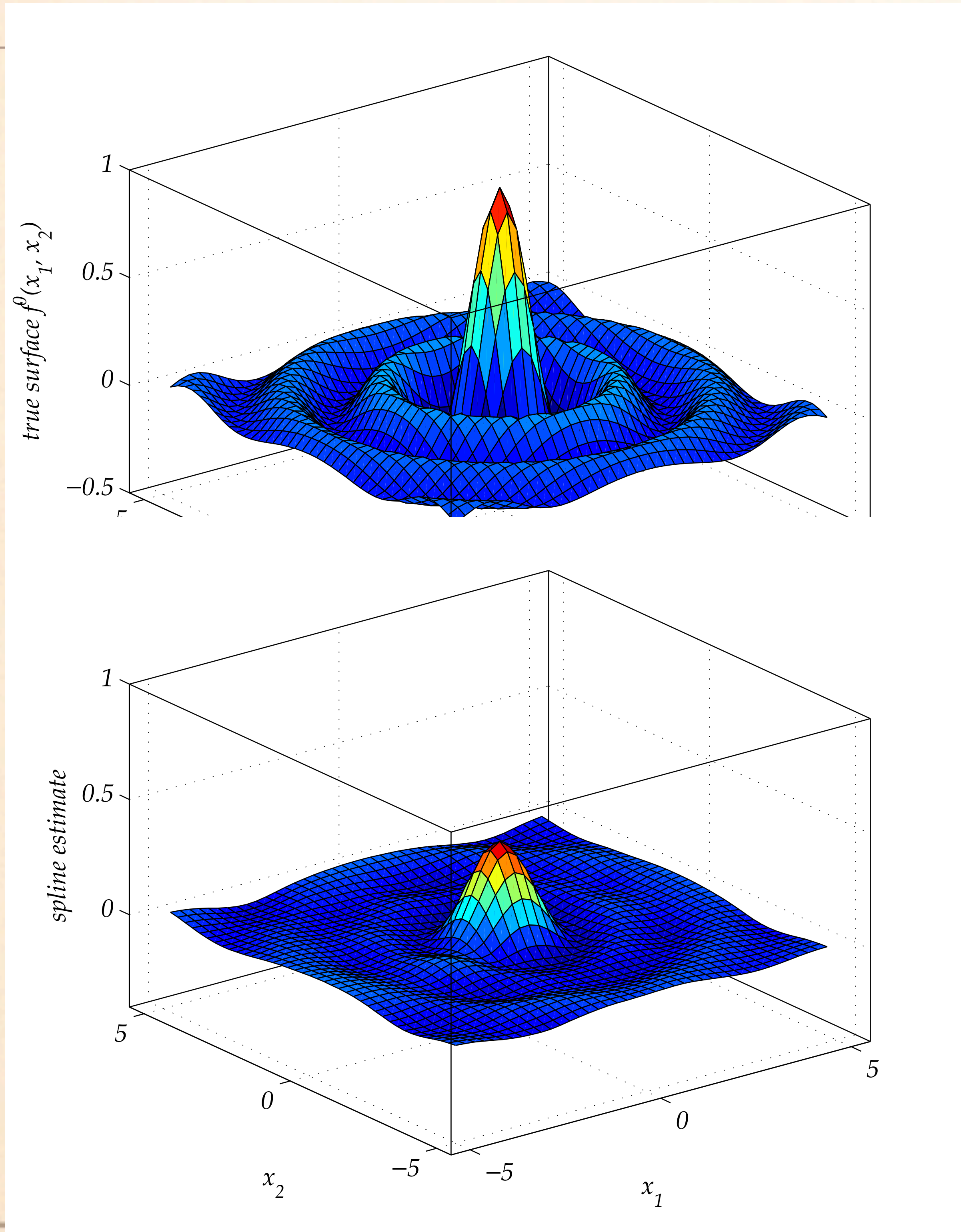


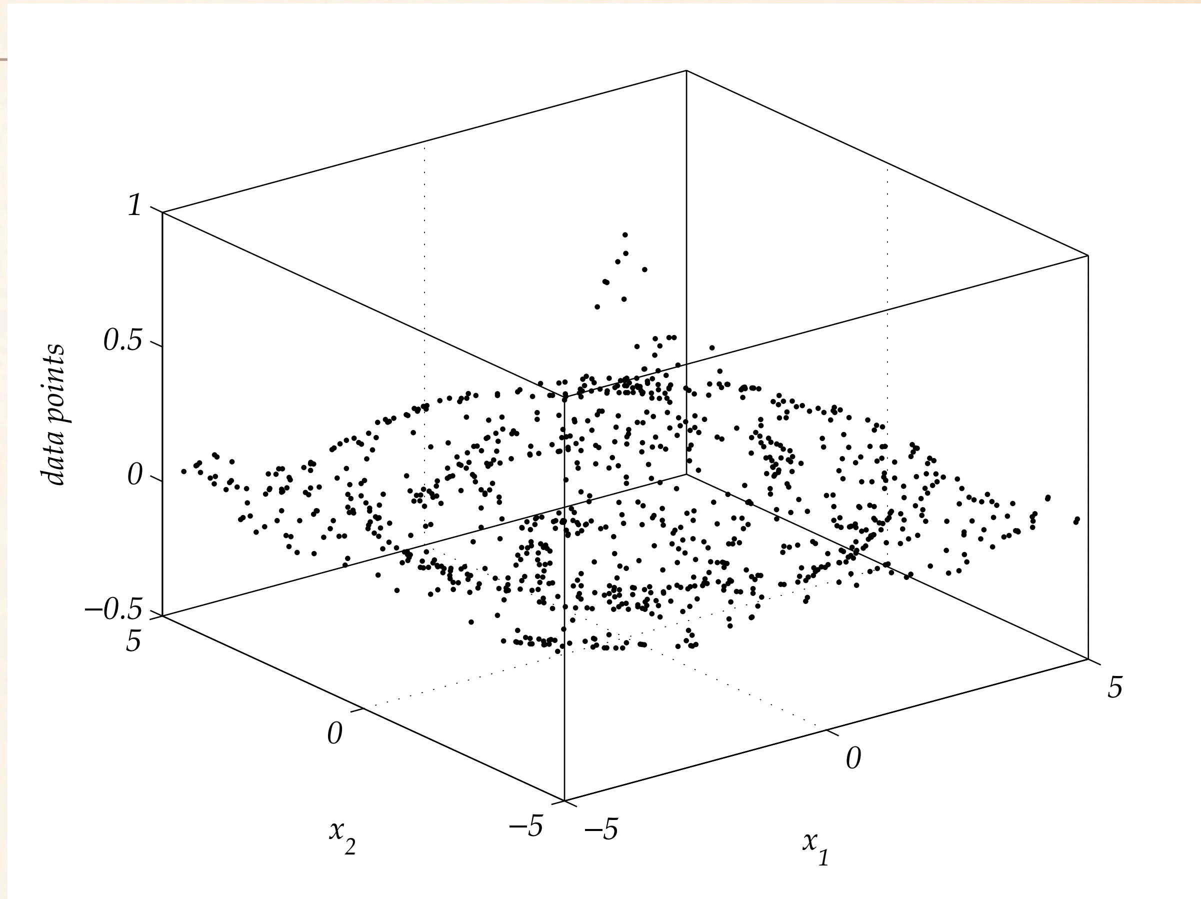
SPLINES & EPI-SPLINES 2



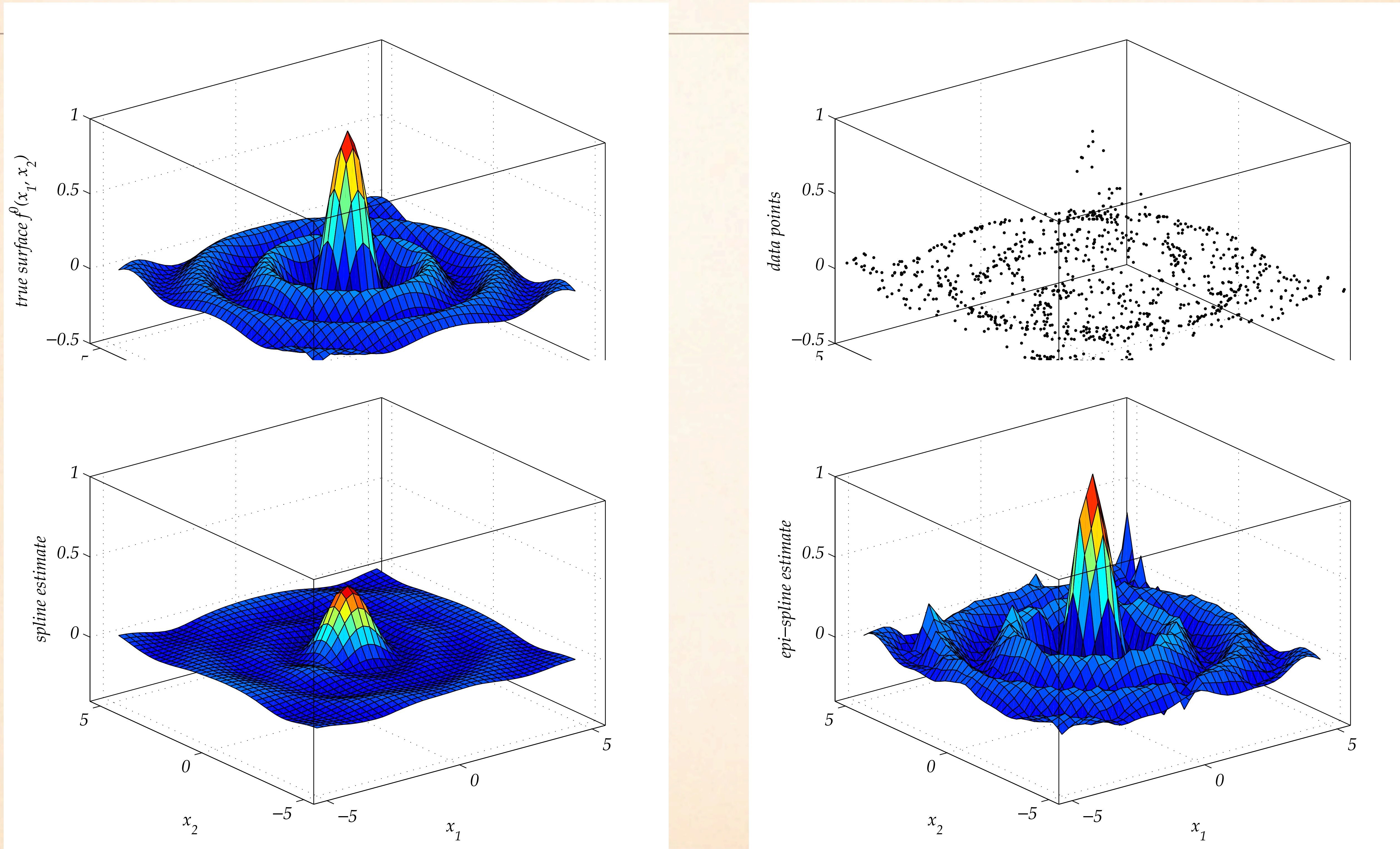


SPLINES & EPI-SPLINES 2

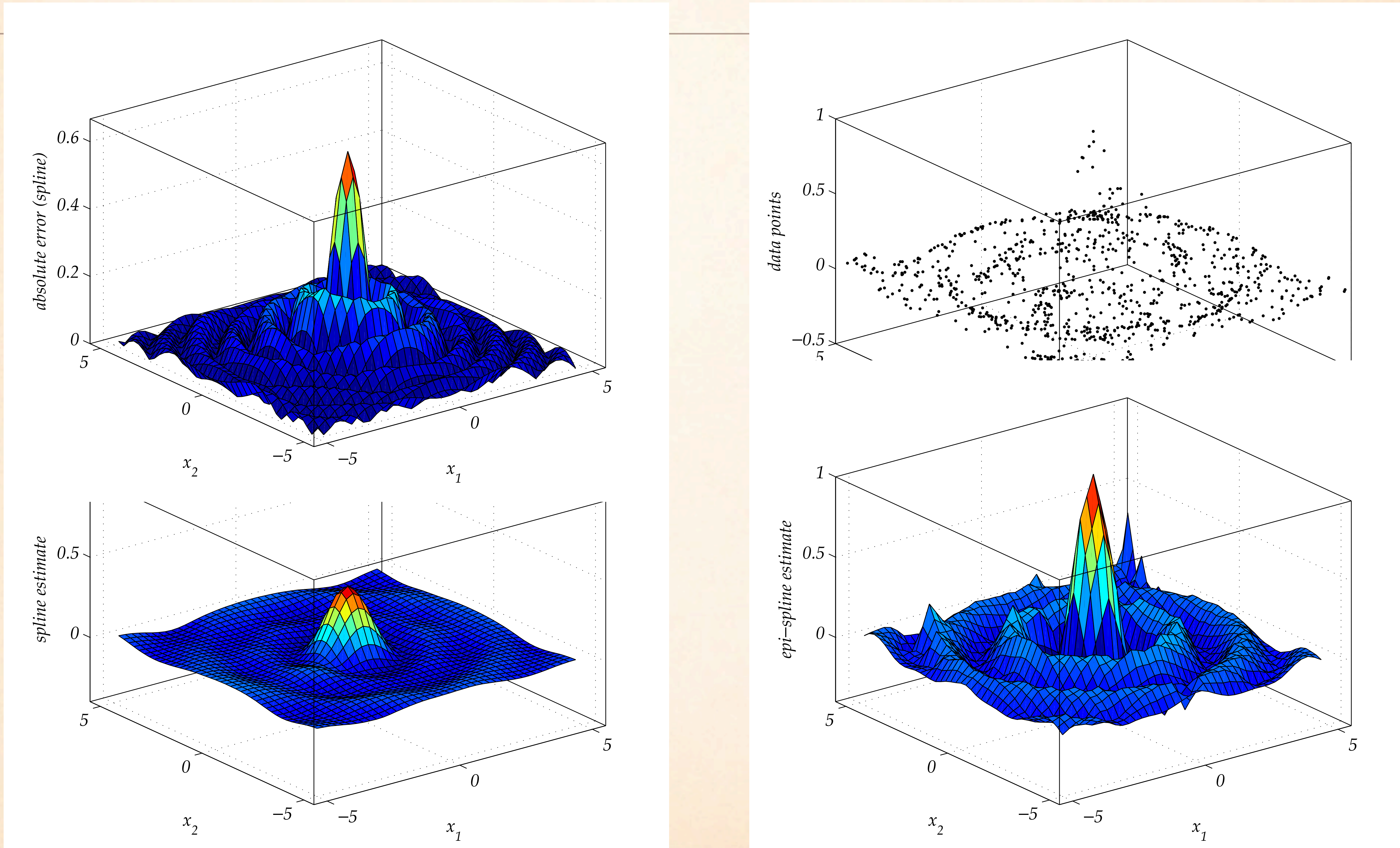




SPLINES & EPI-SPLINES 2



SPLINES & EPI-SPLINES 2



SPLINES & EPI-SPLINES 2