Maximum Likelihood Estimators and LR Statistics for some Non-Regular Problems: Monotone and Convex Functions

Jon A. Wellner
University of Washington

work with Piet Groeneboom, Moulinath Banerjee, & Geurt Jongbloed

Research Day Talk at UW, September, 2000

Email: jaw@stat.washington.edu
http: //www.stat.washington.edu/jaw/jaw.research.html

OUTLINE:

1. **Introduction:**
   regular and irregular statistical problems
2. **Estimating a Monotone Function:**
   - Finite-sample problem.
   - Limit Gaussian problem.
3. **LR Tests for Monotone Functions:**
   - The Limit Gaussian problem.
   - Mouli's theorem.
4. **Estimating a Convex Function:**
   - Finite-sample problem.
   - Limit Gaussian problem.
5. **Further Problems.**

1. Introduction: regular and irregular statistical problems

Regular Problems:

$X_1, \ldots, X_n$ i.i.d. $p(x;\theta)$, $\theta \in \Theta \subset R^d$,
$\theta \mapsto p(\cdot;\theta)$ "smooth". The MLE of θ is

$$\widehat{\theta}_n = \text{argmax}_\theta \sum_{i=1}^{n} \log p(X_i;\theta) \equiv \text{argmax} \log L_n(\theta)\,;$$

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d N_k(0, I(\theta_0)^{-1})$$

where

$$I(\theta_0) = E\left(\dot{l}_\theta(X)\dot{l}_\theta(X)^T\right)$$

is the Fisher Information matrix. The Likelihood Ratio λ_n for testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ is

$$\lambda_n \equiv \frac{\sup_\theta L_n(\theta)}{\sup_{\theta_0} L_n(\theta)} = \frac{L_n(\widehat{\theta}_n)}{L_n(\theta_0)} \equiv LR\,.$$

Then for regular problems,

$$2\log\lambda_n \to_d \chi^2_k$$

"Irregular" Problems:

- Rate of convergence is **not** $n^{1/2}$. Frequently slower!
- Limit distributions are often **not** normal.
- Likelihood ratio statistics are **not** asymptotically chi-square.

Two types of Irregular Problems:

- Estimate a **monotone** function nonparametrically.
- Estimate a **convex** function nonparametrically.

Questions:

1. What are the rates of convergence?
2. What are the asymptotic distributions of the estimators?
3. Do the likelihood ratio statistics still converge in distribution to something free of parameters involved in the problem? What are the limiting distributions?

2. Estimating a Monotone Function

- **Example 1.** Monotone density function on $[0, \infty)$.
  Test $H_0 : f(t_0) = \theta_0$ versus $H_1 : f(t_0) \neq \theta_0$.
- **Example 2.** Interval censoring, current status data.
  Test $H_0 : F(t_0) = \theta_0$ versus $H_1 : F(t_0) \neq \theta_0$.
- **Example 3.** Panel count data.
  Test $H_0 : \Lambda(t_0) = \theta_0$ versus $H_1 : \Lambda(t_0) \neq \theta_0$.
- **Example 4.** Monotone hazard function with right-censored data.
  Test $H_0 : \lambda(t_0) = \theta_0$ versus $H_1 : \lambda(t_0) \neq \theta_0$.
- **Example 5.** Monotone regression function.
  Test $H_0 : r(t_0) = \theta_0$ versus $H_1 : r(t_0) \neq \theta_0$.

Example 2: Current status data

$$X \sim F, \qquad T \sim G\,.$$

We observe $(T, 1\{X \le T\}) \equiv (T, \Delta) \equiv Y$.

$$p_F(t,\delta) = F(t)^{\delta}(1-F(t))^{\delta}\,.$$

Suppose that $Y_i \equiv (T_i, \Delta_i)$ are i.i.d. as (T, Δ).
Test

$$H_0 : F(t_0) = \theta_0 \qquad \text{versus} \qquad H_1 : F(t_0) \neq \theta_0\,.$$

Unconstrained MLE: with $\mathbb{G}_n(t) = n^{-1}\sum_{i=1}^n 1\{T_i \le t\}$,

$$\widehat{\mathbb{F}}_n(t) = \text{left derivative of cumsum diagram at } \mathbb{G}_n^{-1}(t)$$

If $f(t_0) > 0$ and $g(t_0) > 0$, then

$$n^{1/3}(\widehat{\mathbb{F}}_n(t_0) - F(t_0)) \to_d \left\{ \frac{F(t_0)(1-F(t_0))f(t_0)}{2g(t_0)} \right\}^{1/3} 2\mathbb{Z}$$

where, with W a two-sided Brownian motion process starting from 0,

$$\begin{aligned} 2\mathbb{Z} \ &\equiv\ 2\,\text{armin}(W(t)+t^2) \\ &=_d\ \text{slope of greatest convex minorant} \\ &\qquad\qquad \text{of } (W(t)+t^2) \ \text{ at } \ 0\,. \end{aligned}$$

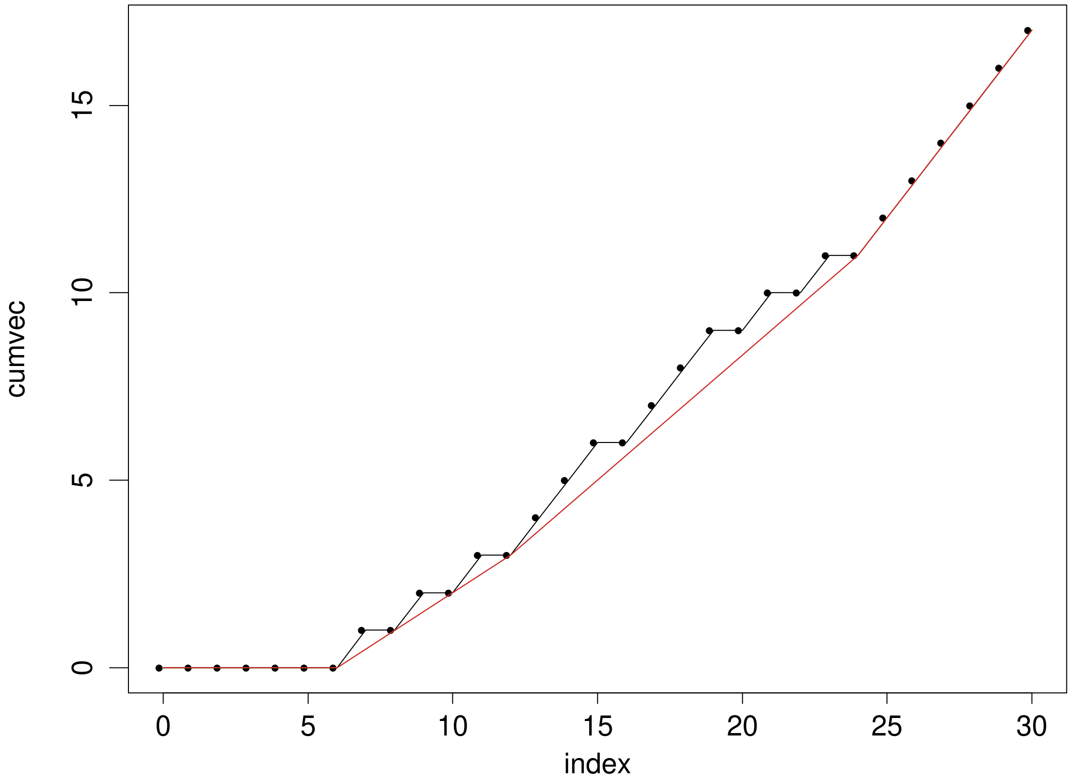

Figure 1: Cumulative sum diagram and Greatest Convex Minorant.

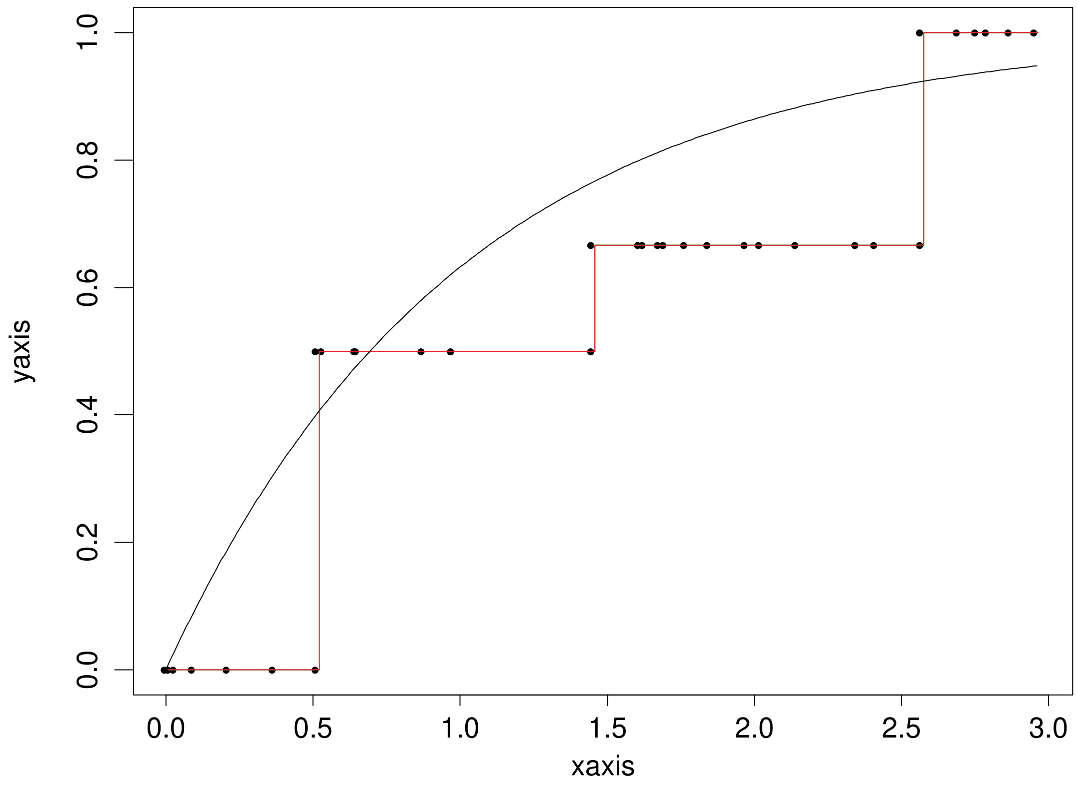

Figure 2: The unconstrained estimator (red).

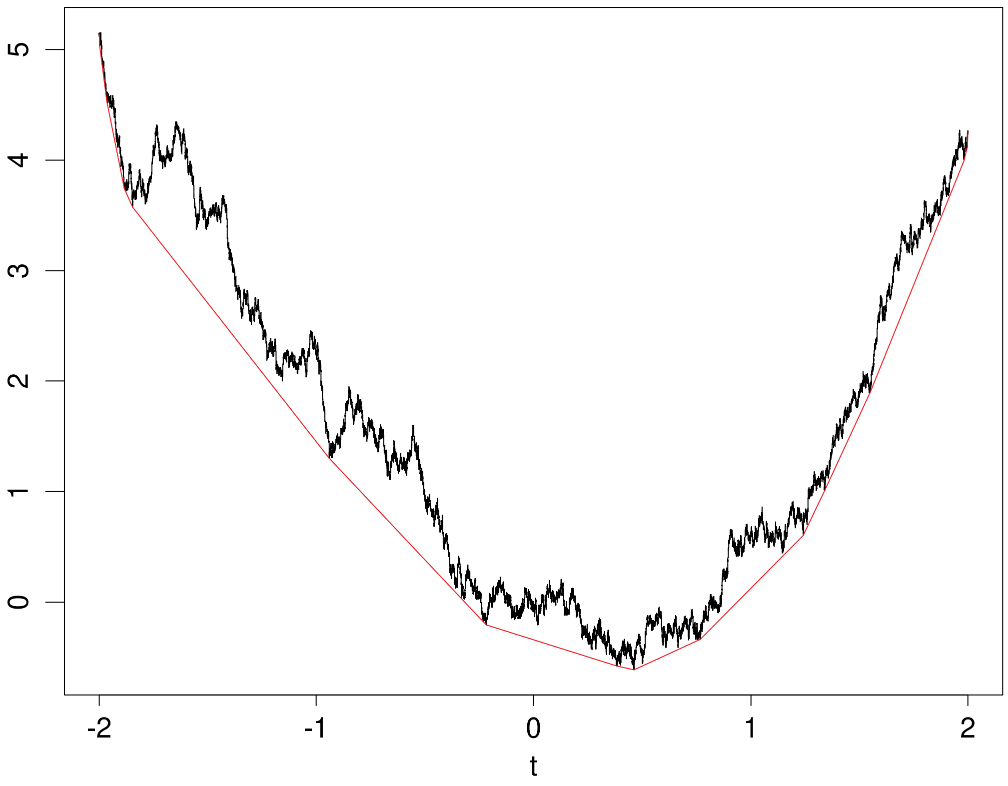

Figure 3: Two-sided Brownian Motion + parabola and GCM

$$X(t) = t^2 + W(t)$$
$$dX(t) = 2t + dW(t)\,;$$

Brownian Scaling Relationships:

$$W(c^2 t) \stackrel{\mathcal{D}}{=} cW(t)\,,$$

so for the slope process $\mathbb{S}_{a,b}$ of

$$X_{a,b}(t) = at^2 + bW(t)\,,$$

$$\mathbb{S}_{a,b}(t) \stackrel{\mathcal{D}}{=} b(a/b)^{1/3}\mathbb{S}_{1,1}((a/b)^{2/3}t) = b(a/b)^{1/3}\mathbb{S}((a/b)^{2/3}t)$$
$$\mathbb{S}_{a,b}(0) \stackrel{\mathcal{D}}{=} b(a/b)^{1/3}\mathbb{S}_{1,1}(0) = b(a/b)^{1/3}\mathbb{S}(0)$$

Groeneboom (1985), (1989) , Groeneboom and Wellner (2000)

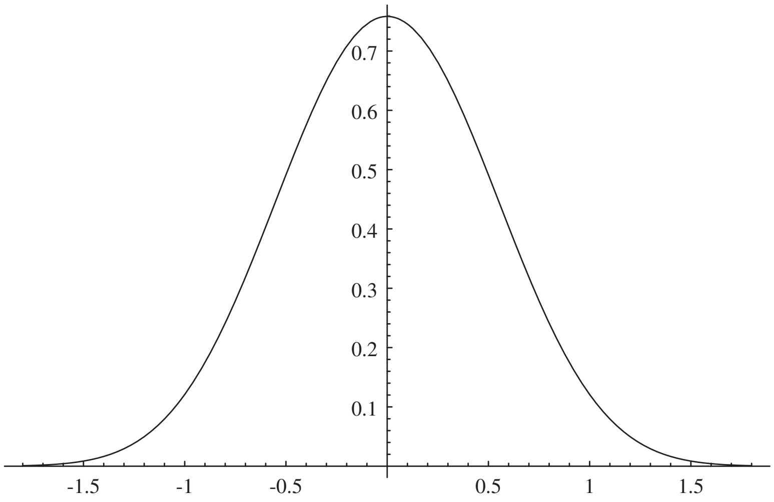

Figure 4: Density function of Z, f_Z.

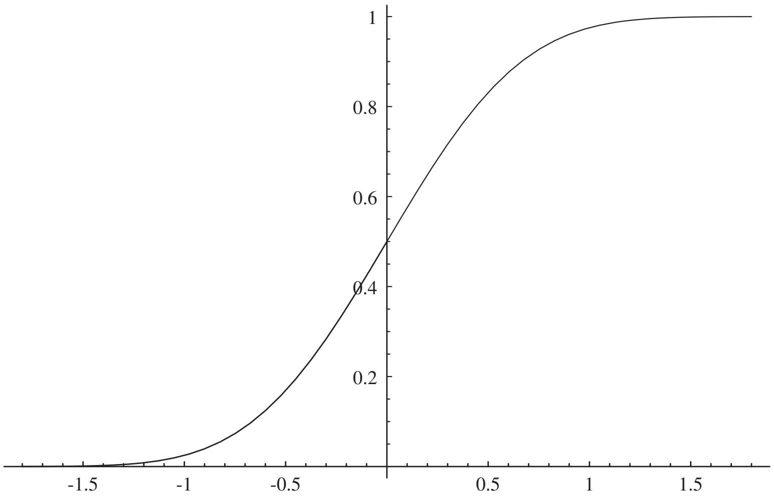

Figure 5: Distribution function of Z, F_Z.

3. LR Tests for Monotone Functions:

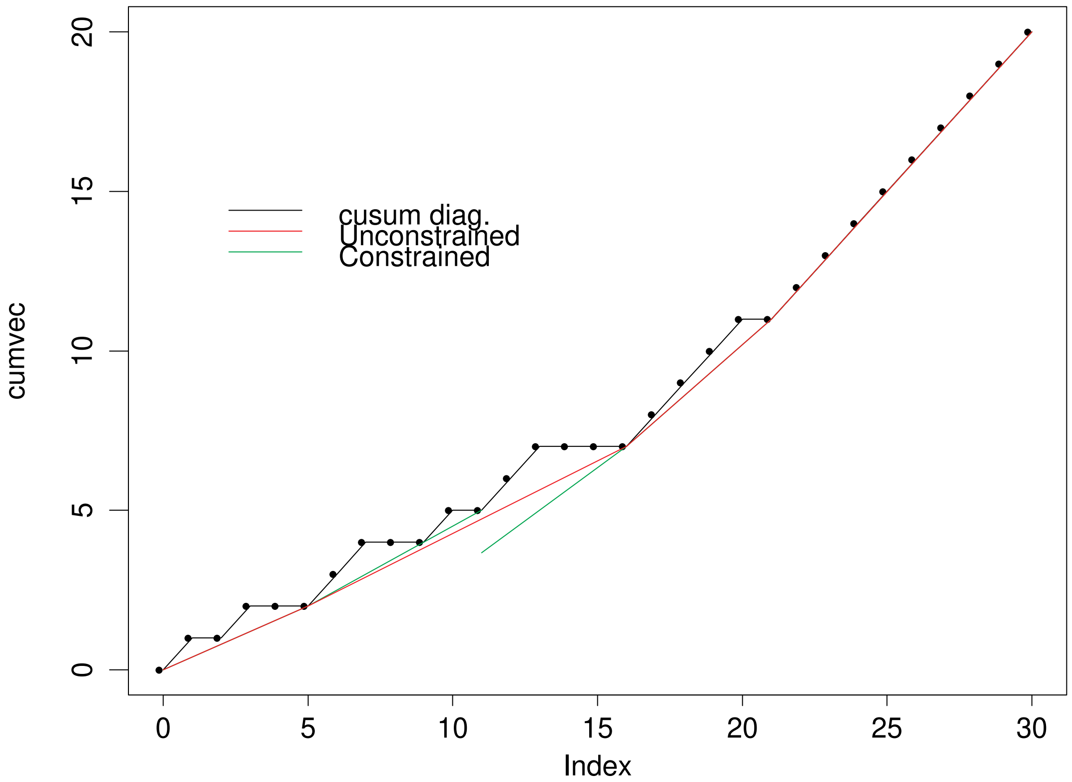

Figure 6: Cumulative sum diagram with left and right Greatest Convex Minorants.

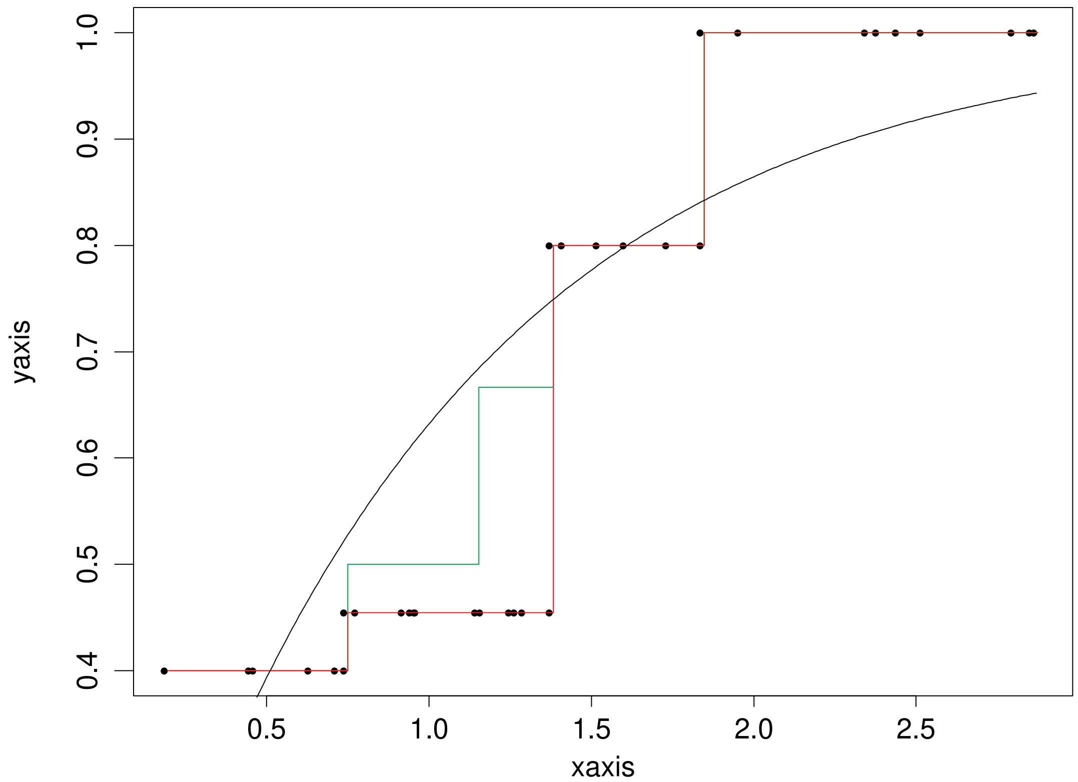

Figure 7: The constrained estimator (green).

The Likelihood Ratio Statistic:

$$\log L_n(F, Y_1, \ldots, Y_n) = \sum_{i=1}^{n} \{\Delta_i \log F(T_i) + (1 - \Delta_i)(1 - F(T_i)\} \, .$$

$$\begin{aligned} \text{LR statistic} \equiv \lambda_n(\theta_0) \equiv \lambda_n &= \frac{\sup_F L_n(F, Y_1, \ldots, Y_n)}{\sup_{F(t_0)=\theta_0} L_n(F, Y_1, \ldots, Y_n)} \\ &= \frac{L_n(\widehat{\mathbb{F}}_n, Y_1, \ldots, Y_n)}{L_n(\widehat{\mathbb{F}}_n^0, Y_1, \ldots, Y_n)} \end{aligned}$$

where $\widehat{\mathbb{F}}_n^0$ is the MLE of F subject to the constraint $\widehat{\mathbb{F}}_n^0(t_0) = \theta_0$.

Question:

$$2 \log \lambda_n \to_d \text{something?}$$

Answer: Yes! Mouli's theorem:

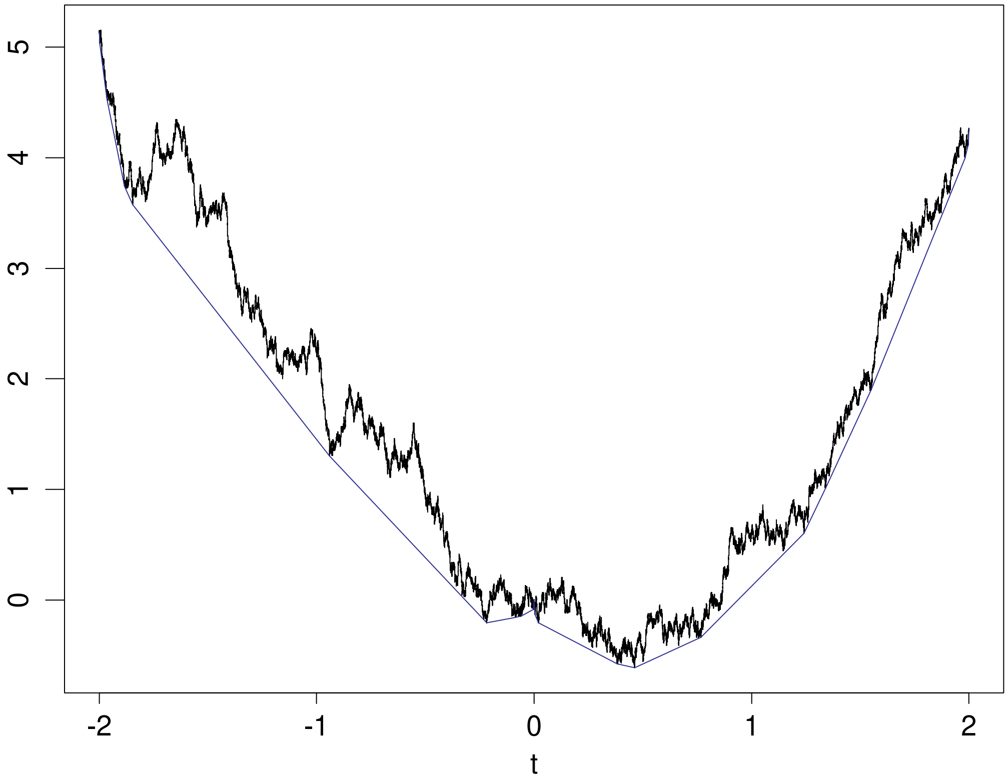

Figure 8: Left and Right GCM's of W + parabola

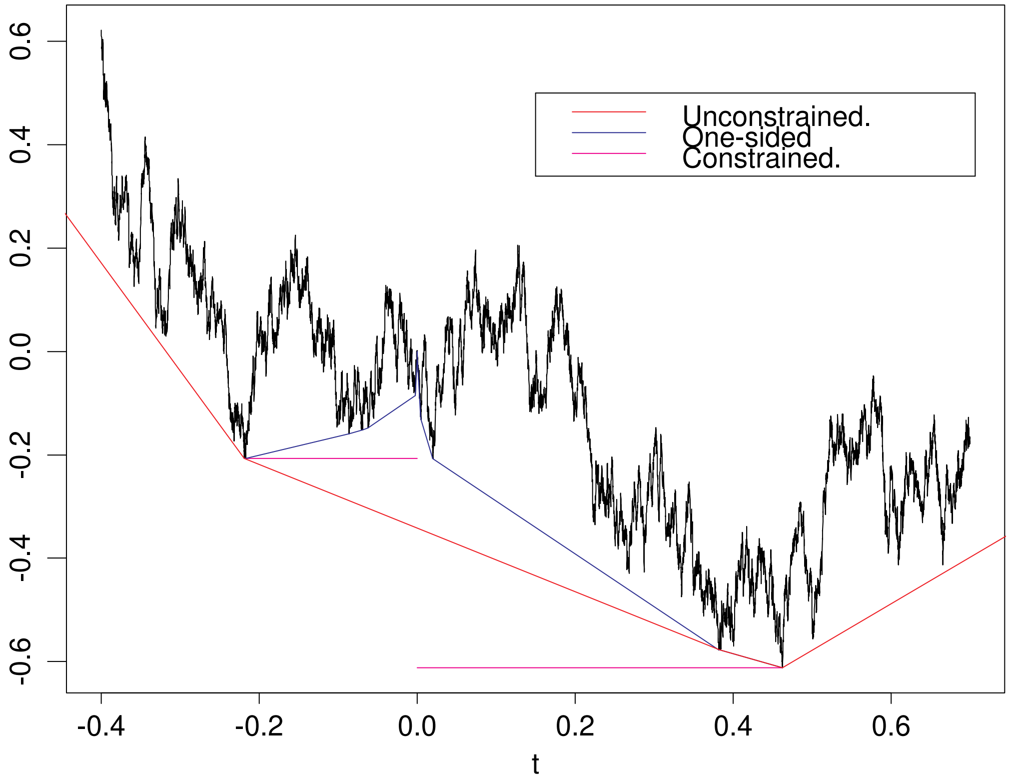

Figure 9: Global, Left, and Right GCM's: a close-up view

Theorem. (Banerjee and Wellner, 2000). Suppose that F and G have densities f and g which are strictly positive and continuous in a neighborhood in a neighborhood of t_0. Suppose that F satisfies the null hypothesis $H_0 : F(t_0) = \theta_0$. Then

$$2 \log \lambda_n \to_d \mathbb{D} = \int \{(\mathbb{S}(z))^2 - (\mathbb{S}^0(z))^2\} dz \,,$$

and the distribution of $\mathbb{D}$ is **universal** (free of parameters).

Confidence Interval? Invert the test:

$$\{\theta : 2 \log \lambda_n(\theta) \le d_\alpha\} \,.$$

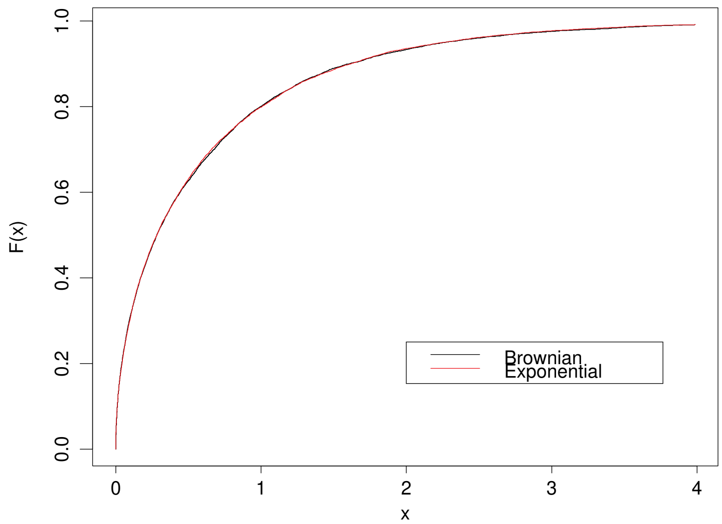

Figure 10: Empirical Distribution of D, 10^4 replications

4. Estimating a Convex Density Function.

Theorem. (Groeneboom, Jongbloed, and Wellner (2000a), (2000b))

$$n^{2/5}(\widehat{f}_n(t) - f(t)) \to \left(\frac{f^2(t)f''(t)}{24}\right)^{1/5} \mathbb{Z}_{con}$$

where $\mathbb{Z}_{con} = H''(0)$ for a process H defined as follows:

$$\begin{aligned} dX(t) &= 12t^2 + dW(t) \\ &= \text{"canonical convex function"} \\ &\qquad\qquad + \text{Gaussian white noise}\,; \\ X(t) &= 4t^3 + W(t)\,; \\ Y(t) &= t^4 + \int_0^t W(s)ds\,. \end{aligned}$$

Theorem. There exists an almost surely uniquely defined random continuous function H satisfying the following conditions:

(i) The function H is everywhere above the function Y:

(1) $H(t) \geq Y(t),$ for each $t \in R$.

(ii) H has a convex second derivative.

(iii) The function H satisfies

(2) $\displaystyle\int_R \{H(t) - Y(t)\}\, dH^{(3)}(t) = 0.$

5. Further Problems.

A. Can we find the distribution of $\mathbb{D}$ analytically?

B. How can we prove that the same limit $\mathbb{D}$ arises as the limit distribution for the likelihood ratio test for a large class of such problems involving monotone functions?

C. What is the appropriate contiguity theory? What is the limit distribution of the likelihood ratio statistic under local alternatives?

D. What happens if we constrain at $k > 1$ points?

E. Can we use a union-intersection test to obtain confidence bands for the whole monotone function F?

F. Distribution of $\mathbb{Z}_{con}$ via Monte-Carlo?

G. Can we find the distribution of $\mathbb{Z}_{con}$ analytically?

H. Can we extend the results for the LR statistic from the monotone case to the convex case?

Selected References:

- Banerjee, M. and Wellner, J. A. (2000). Likelihood ratio tests for monotone functions. *Technical Report* **377**, Department of Statistics, University of Washington.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2000a). A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion $+t^4$. *Technical Report* **369**, Department of Statistics, University of Washington.
- Groeneboom, P., Jongbloed, G., and Wellner, J. A. (2000b). Estimation of convex functions: characterizations and asymptotic theory. *Technical Report* **372**, Department of Statistics, University of Washington.
- Groeneboom, P. and Wellner, J. A. (2000). Computing Chernoff's distribution. *J. Computational and Graphical Statistics*, to appear.
- Wellner, J. A. (2000). Gaussian white noise models: some results for monotone functions. *Technical Report* **??**, Department of Statistics, University of Washington.