Semiparametric Data Fusion

Benjamin Kedem

Department of Mathematics & Inst. for Systems Research
University of Maryland, College Park

"Give me a place to stand and rest my lever on, and I can move the Earth",
(Archimedes, 287-212 B.C.)

Jerry North Symposium, June 8-9, 2009

Outline

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- k-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- k-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- *k*-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- *k*-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- k-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- *k*-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

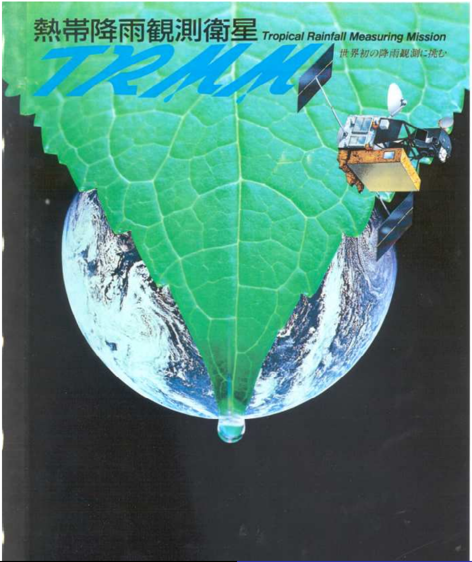
Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- *k*-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

Anything in common?

- Satellite sensors and their ground truth counterpart.
- An array of sensors sensing a common target.
- Multiple filtering of a stationary signal.
- Multivariate autoregression with Gaussian noise.
- Classical analysis of variance with normal data.
- k-parameter exponential families.
- Clinical case control systems.
- Weighted distributions and biased sampling.

TRMM sensors: Likely distortions of ground truth



wooden_bucket.jpg (JPEG Image, 167x218 pixels) http://upload.wikimedia.org/wi

Reference: Ground truth

An array of sensors sensing a common target

We don't have to go very far...

- Consider a "system" of two ears, and a whisper coming from the right.
- The right ear is the more reliable sensor. Think of it as a reference.
- The left ear is a distortion of the reference.

Multiple filtering of a signal

$$
\begin{aligned}
f_1(\omega) &= |H_1(\omega)|^2 f(\omega) \\
&\cdot \\
&\cdot \\
&\cdot \\
f_q(\omega) &= |H_q(\omega)|^2 f(\omega)
\end{aligned} \tag{1}
$$

That is, q “distortions” or multiple “tilting” of the same *reference* spectral density f.

Linear system with Gaussian noise

$$\boldsymbol{X}_t = \boldsymbol{\alpha} \boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon} = (\epsilon_{1t}, ..., \epsilon_{qt}, \epsilon_{mt})'$$

$\epsilon_{jt} \sim g_j \sim N(0, \sigma_j^2)$, $j = 1, ..., q, m$. Choose $g_m \equiv g$. We get many distortions of the same *reference* g:

$$g_1(x) = e^{\alpha_1 + \beta_1 x^2} g(x)$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$g_q(x) = e^{\alpha_q + \beta_q x^2} g(x)$$

Analysis of variance

Consider the classical one-way ANOVA with $m = q + 1$ independent normal random samples:

$$x_{11}, \quad \ldots \quad , x_{1n_1} \sim g_1(x)$$

$$\vdots$$

$$x_{q1}, \quad \ldots \quad , x_{qn_q} \sim g_q(x)$$

$$x_{m1}, \quad \ldots \quad , x_{mn_m} \sim g_m(x)$$

$$g_j(x) \sim \mathrm{N}(\mu_j, \sigma^2), \qquad j = 1, ..., m.$$

Then, holding $g_m(x) \equiv g(x)$ as a reference:

$$\begin{aligned} g_1(x) &= \exp(\alpha_1 + \beta_1 x) g(x) \\ &\vdots \\ g_q(x) &= \exp(\alpha_q + \beta_q x) g(x) \end{aligned}$$

$$\alpha_j = \frac{\mu_m^2 - \mu_j^{\mathbf{2}}}{2\,\sigma^2}, \qquad \beta_j = \frac{\mu_j - \mu_{\mathbf{m}}}{\sigma^2}, \qquad j = 1, \ldots, q$$

k-parameter exponential families

$$g(x,\boldsymbol{\theta}) = d(\boldsymbol{\theta})S(x)\exp\left\{\sum_{i=1}^{k} c_i(\boldsymbol{\theta})T_i(x)\right\}$$

Let $g_j(x) \equiv g(x,\boldsymbol{\theta}_j)$, $g(x) \equiv g(x,\boldsymbol{\theta}_m)$.
Again, q distortions of a reference $g(x)$:

$$\begin{aligned} g_1(x) &= \exp\{\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{h}(x)\}g(x) \\ &\vdots \\ g_q(x) &= \exp\{\alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(x)\}g(x) \end{aligned}$$

Case-control: Multinomial logistic regression (Prentice and Pyke 1979).

- RV y s.t. $P(y=j) = \pi_j, \quad \sum_{j=1}^{m} \pi_j = 1.$
- Assume: For $j = 1, ..., m$, and *any* $h(x)$,

$$P(y=j|x) = \frac{\exp(\alpha_j^* + \beta_j h(x))}{1 + \sum_{k=1}^{q} \exp(\alpha_k^* + \beta_k h(x))}$$

- Define: $f(x|y=j) = g_j(x), \quad j = 1, ..., m$

Then with $\alpha_j = \alpha_j^* + \log[\pi_m/\pi_j]$, $j = 1, ..., q$, and $g_m \equiv g$,

$$g_j(x) = \exp(\alpha_j + \beta_j h(x)) g(x), \; j = 1, ..., q.$$

Weighted Distributions

Rao(1965) unified the concept of weighted distributions of the form:

$$p_w(x;\theta,\alpha) = \frac{W(x;\alpha)p(x;\theta)}{E[W(X;\alpha)]}$$

where $W(x;\alpha)$ is known. This is an example of "tilting" of a reference distribution. By changing the weight $W(x;\alpha)$ we obtain different distortions of the same reference $p(x;\theta)$.

Biased sampling

- **Length-biased Sampling:** Vardi (1982) introduced

$$F(y) = 1/\mu \int_0^y x dG(x), \; y \geq 0,$$

where $\mu = \int_0^\infty x dG(x) < \infty$. This is a tilt model.

- **Biased Sampling/Selection Bias:** Vardi (1985), and Gill, Vardi, Wellner (1988) considered the more general biased sampling model

$$F(y) = W(G)^{-1} \int_{-\infty}^y w(x) dG(x),$$

where $w(x)$ is known and $W(G) = \int_{-\infty}^\infty w(x) dG(x)$. This is a tilt model. F, G are obtained by NPMLE.

Biased sampling

- **Length-biased Sampling:** Vardi (1982) introduced

$$F(y) = 1/\mu \int_0^y x dG(x), \; y \geq 0,$$

  where $\mu = \int_0^\infty x dG(x) < \infty$. This is a tilt model.

- **Biased Sampling/Selection Bias:** Vardi (1985), and Gill, Vardi, Wellner (1988) considered the more general biased sampling model

$$F(y) = W(G)^{-1} \int_{-\infty}^y w(x) dG(x),$$

  where $w(x)$ is known and $W(G) = \int_{-\infty}^\infty w(x) dG(x)$. This is a tilt model. F, G are obtained by NPMLE.

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$
\begin{aligned}
g_1 &= w_1 g \\
&\cdot \\
&\cdot \\
&\cdot \\
g_q &= w_q g
\end{aligned}
$$

- How can we take advantage of this?
- Can the distorted information be of use?
- Can the distorted and reference information be fused as to improve the reference information?
- In other words, can the "bad" improve the "good"?

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$\begin{aligned} g_1 &= w_1 g \\ &\cdot \\ &\cdot \\ &\cdot \\ g_q &= w_q g \end{aligned}$$

- How can we take advantage of this?
- Can the distorted information be of use?
- Can the distorted and reference information be fused as to improve the reference information?
- In other words, can the “bad” improve the “good”?

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$\begin{aligned} g_1 &= w_1 g \\ &\cdot \\ &\cdot \\ &\cdot \\ g_q &= w_q g \end{aligned}$$

- How can we take advantage of this?
- Can the distorted information be of use?
- Can the distorted and reference information be fused as to improve the reference information?
- In other words, can the “bad” improve the “good”?

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$
\begin{aligned}
g_1 &= w_1 g \\
&\cdot \\
&\cdot \\
&\cdot \\
g_q &= w_q g
\end{aligned}
$$

- How can we take advantage of this?
- Can the distorted information be of use?
- Can the distorted and reference information be fused as to improve the reference information?
- In other words, can the "bad" improve the "good"?

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$\begin{aligned} g_1 &= w_1 g \\ &\cdot \\ &\cdot \\ &\cdot \\ g_q &= w_q g \end{aligned}$$

- How can we take advantage of this?
- Can the distorted information be of use?
- Can the distorted and reference information be fused as to improve the reference information?
- In other words, can the “bad” improve the “good”?

- The relationship between a reference and its distortions opens the door to a useful general statistical approach based on *fused information* from many sources.

The above observations suggest the following general semiparametric problem.

- Multiple data sources: $\boldsymbol{x}_1, ..., \boldsymbol{x}_q, \boldsymbol{x}_m$.
- Data fusion: $\mathbf{t} = (t_1, ..., t_n)' \equiv (\mathbf{x}_1', ..., \mathbf{x}_q', \mathbf{x}_m')'$.
- Fused data length: $n \equiv n_1 + \cdots + n_q + n_m$.
- Assume: $\boldsymbol{x}_j \sim g_j(x), \quad j = 1, ..., q, m$.
- Reference pdf: $g_m(x) = g(x)$.
- Tilting: $g_j(x) = \exp(\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(x)) g(x), \qquad j = 1, ..., q$.

Problem: Use the fused data $\mathbf{t} = (t_1, ..., t_n)'$ to

1. Estimate the reference pdf $g(x)$ and cdf $G(x)$.

2. Estimate $\boldsymbol{\alpha}=(\alpha_1, ..., \alpha_q)'$, $\boldsymbol{\beta}=(\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_q')'$.

3. Test hypotheses about the $\boldsymbol{\beta}_i$, and in particular test distribution equality, H_0: $\boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_{\mathrm{q}} = 0$

Outline

Estimation

Follow Qin and Zhang (1997), FKQS (2001).

MLE of $G(x)$ can be obtained by maximizing the likelihood over the class of step cdf's with jumps at the observed values $t_1, ..., t_n$. Accordingly, if $p_i = dG(t_i)$, $i = 1, .., n$, the empirical likelihood becomes,

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp(\ \alpha_1 + \ \boldsymbol{\beta}_1' \boldsymbol{h}(x_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\ \alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(x_{qj}))$$

1. Get p_i

Fix α, β. Maximize $\prod_{i=1}^{n} p_i$ subject to the m constraints:

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i[w_j(t_i) - 1] = 0, \qquad j = 1, ..., q,$$

where

$$w_j(t_i) = \exp(\ \alpha_j + \ \beta_j' \boldsymbol{h}(t_i)), \ j = 1, ..., q.$$

Use Lagrange multipliers $\lambda_0 = n, \qquad \lambda_j = \nu_j n$. Then

$$(\star) \qquad p_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)}$$

where

$$(\star) \qquad \rho_j = n_j / n_m, \qquad j = 1, ..., q.$$

2. Estimate $\boldsymbol{\alpha}, \boldsymbol{\beta}$

Profile log-likelihood up to a constant as a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ only:

$$l = \sum_{j=1}^{n_1} [\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{h}(x_{1j})] + \cdots + \sum_{j=1}^{n_q} [\alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(x_{qj})]$$
$$- \sum_{i=1}^{n} \log[1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)]$$

Score equations for $j = 1, ..., q$:

$$\frac{\partial l}{\partial \alpha_j} = -\sum_{i=1}^{n} \frac{\rho_j w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + n_j = 0$$

$$\frac{\partial l}{\partial \beta_j} = -\sum_{i=1}^{n} \frac{\rho_j h(t_i) w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + \sum_{i=1}^{n_j} h(x_{ji}) = 0$$

3. Estimate $g(x), G(x)$

The solution of the score equations gives the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}$, and consequently by substitution also $\hat{p}_i$. Thus,

$$\hat{p}_{\mathbf{i}} = \frac{1}{n_m} \cdot \frac{1}{1 + \sum_{j=1}^{q} \rho_j \exp(\hat{\alpha}_j + \hat{\beta}_j' \boldsymbol{h}(t_i))}.$$

Therefore,

$$\hat{g}(x) = Kernel(\hat{p}_{\mathbf{i}})$$

and

$$\hat{G}(t) = \sum_{i=1}^{n} I(t_i \leq t) \hat{p}_{\mathbf{i}}$$

Everything is estimated from everything

The reference $G(x)$ and all the parameters, and hence all the tilted distributions, are estimated from the entire fused data $\boldsymbol{t}$. Thus $G(x)$ is estimated from the fused data $\boldsymbol{t}$ and not just from the reference sample $\boldsymbol{x}_m$, and β_1 is estimated from $\boldsymbol{t}$ and not just from $\boldsymbol{x}_1$, etc. This **borrowing of strength** leads to more precise estimation. We shall quantify this in kernel density estimation.

Some asymptotic results

Assumptions

- The second moments of $h(t)$ with respect to each distribution are finite,

$$\int h^2(t) w_j(t) dG(t) < \infty,$$

$j = 1, \ldots, m.$

- The relative sample sizes $\rho_j = n_j/n_m$ are finite and remain fixed as the total sample size $\sum_{j=1}^m n_j = n \to \infty$, $j = 1, \ldots, m.$

Some asymptotic results

Assumptions

- The second moments of $h(t)$ with respect to each distribution are finite,

$$\int h^2(t) w_j(t) dG(t) < \infty,$$

$j = 1, \ldots, m.$

- The relative sample sizes $\rho_j = n_j / n_m$ are finite and remain fixed as the total sample size $\sum_{j=1}^{m} n_j = n \to \infty$, $j = 1, \ldots, m$.

Fact (QZ (1997), FKQS (2001)): Assume

$$g_j(x) = \exp(\alpha_j + \beta_j h(x))g(x), \qquad j = 1, \ldots, q,$$

with true parameters $\boldsymbol{\alpha}_0 = (\alpha_{10}, \ldots, \alpha_{q0})'$, $\quad \boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{10}, \ldots, \boldsymbol{\beta}_{q0})'$. Then under regularity conditions the MLE's $\hat{\alpha}, \hat{\beta}$ are strongly consistent and asymptotically normal,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \Rightarrow \mathrm{N}\left(\mathbf{0}, \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}\right).$$

$\boldsymbol{V}, \boldsymbol{S}$ are defined below. We sometimes write $\boldsymbol{\Sigma} = \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}$.

$$\boldsymbol{V} \equiv Var\left[\frac{1}{\sqrt{n}}\nabla l(\alpha,\beta)\right], \quad \boldsymbol{S} \equiv \lim_{n\to\infty} -\frac{1}{n}\nabla\nabla' l(\alpha,\beta)$$

where $\boldsymbol{V}$,$\boldsymbol{S}$ are $q(1+p)\times q(1+p)$ matrices.

Remark:

The entries in $\boldsymbol{S}$ are obtained by a repeated application of the facts

$$\int dG(t) = 1, \quad \int w_j(t)dG(t) = 1, \quad j = 1, ..., q.$$

Remark:

It should be noted that due to profiling, the matrix $\boldsymbol{S}$ is not the usual information matrix but it plays a similar role.

Define $\rho_m \equiv 1, \; w_m(t) \equiv 1$,

$$E_j[\boldsymbol{h}(t)] \equiv \int \boldsymbol{h}(t) w_j(t) dG(t)$$

and,

$$\begin{aligned}
A_0(j, j') &\equiv \int \frac{w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \\
\boldsymbol{A}_1(j, j') &\equiv \int \frac{\boldsymbol{h}(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)} \\
\boldsymbol{A}_2(j, j') &\equiv \int \frac{\boldsymbol{h}(t) \boldsymbol{h}'(t) w_j(t) w_{j'}(t) dG(t)}{1 + \sum_{k=1}^{q} \rho_k w_k(t)}
\end{aligned}$$

for $j, j' = 1, ..., q$.

The entries in $\boldsymbol{V}$:

$$
\begin{aligned}
\frac{1}{n} Var\left(\frac{\partial l}{\partial \alpha_j}\right) &= \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j) - \sum_{r=1}^m \rho_r A_0^2(j,r)\} \\
\frac{1}{n} Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \alpha_{j'}}\right) &= \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j') \\
&- \sum_{r=1}^m \rho_r A_0(j,r) A_0(j',r)\} \\
\frac{1}{n} Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_j}\right) &= \frac{\rho_j^2}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j) E_j[\boldsymbol{h}'(t)] \\
&- \sum_{r=1}^m \rho_r A_0(j,r) \boldsymbol{A}_1'(j,r)\} \\
\frac{1}{n} Cov\left(\frac{\partial l}{\partial \alpha_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) &= \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^q \rho_k}\{A_0(j,j') E_{j'}[\boldsymbol{h}'(t)] \\
&- \sum_{r=1}^m \rho_r A_0(j,r) \boldsymbol{A}_1'(j',r)\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n} Cov\left(\frac{\partial l}{\partial \boldsymbol{\beta}_j}, \frac{\partial l}{\partial \boldsymbol{\beta}_{j'}}\right) &= \frac{\rho_j \rho_{j'}}{1+\sum_{k=1}^{q} \rho_k}\{-\boldsymbol{A}_2(j,j') + E_j[\boldsymbol{h}(t)]\boldsymbol{A}_1'(j,j') \\
&+ \boldsymbol{A}_1(j,j')E_{j'}[\boldsymbol{h}'(t)] \\
&- \sum_{r=1}^{m} \rho_r \boldsymbol{A}_1(j,r)\boldsymbol{A}_1'(j',r)\} \\
&+ \frac{1}{n}\sum_{i=1}^{n_j}\sum_{k=1}^{n_{j'}} Cov[\boldsymbol{h}(\epsilon_{ji}), \boldsymbol{h}(\epsilon_{j'k})]
\end{aligned}
$$

The last term is 0 for $j \neq j'$ and $(n_j/n)Var[\boldsymbol{h}(\epsilon_{j1})]$ for $j = j'$.

The entries in $\boldsymbol{S}$:

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j^2} \rightarrow \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \alpha_{j'}} \rightarrow \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \partial \boldsymbol{\beta}_j'} \rightarrow \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \alpha_j \partial \boldsymbol{\beta}_{j'}'} \rightarrow \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_j'} \rightarrow \frac{\rho_j}{1+\sum_{k=1}^q \rho_k}\int \frac{[1+\sum_{k\neq j}^q \rho_k w_k(t)]w_j(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

$$-\frac{1}{n}\frac{\partial^2 l}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_{j'}'} \rightarrow \frac{-\rho_j\rho_{j'}}{1+\sum_{k=1}^q \rho_k}\int \frac{w_j(t)w_{j'}(t)\boldsymbol{h}(t)\boldsymbol{h}'(t)}{1+\sum_{k=1}^q \rho_k w_k(t)}dG(t)$$

Fact (G. Lu (2007)): The process $\sqrt{n}(\hat{G}(t) - G(t))$ converges weakly to a zero-mean Gaussian process in $D[-\infty, \infty]$, with covariance matrix given by

$$\begin{aligned} &\mathrm{Cov}\{\sqrt{n}(\hat{G}(t) - G(t)), \sqrt{n}(\hat{G}(s) - G(s))\} = \\ &\sum_{k=0}^{m} \rho_k \Big(G(t \wedge s) - G(t)G(s) - \sum_{j=1}^{m} \rho_j A_j(t \wedge s) \Big) \\ &+ \Big(\bar{A}'(s)\boldsymbol{\rho}, \bar{B}'(s)(\boldsymbol{\rho} \otimes \mathbf{1}_p) \Big) S^{-1} \begin{pmatrix} \boldsymbol{\rho}\bar{A}(t) \\ (\boldsymbol{\rho} \otimes \mathbf{1}_p)\bar{B}(t) \end{pmatrix}. \end{aligned} \tag{2}$$

where $\boldsymbol{\rho} = \mathrm{diag}\{\rho_1, \ldots, \rho_m\}$, ρ_j's are sample fractions, $\bar{A}$ and $\bar{B}$ are vectors of first and second weighted moments of the distortion function $\boldsymbol{h}$ with respect to different samples, and $\mathbf{1}_p = (1, ..., 1)'$. (Note: here $\boldsymbol{h}$ is vector-valued).

Hypothesis testing

Under $\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0}$, all the moments are taken with respect to the reference g.

Define a $q \times q$ matrix $\mathbf{A}_{11}$ whose jth diagonal element is

$$\frac{\rho_j[1 + \sum_{k \neq j}^{q} \rho_k]}{[1 + \sum_{k=1}^{q} \rho_k]^2}.$$

For $j \neq j'$, the jj' element is

$$\frac{-\rho_j \rho_{j'}}{[1 + \sum_{k=1}^{q} \rho_k]^2}.$$

The elements are bounded by 1 and the matrix is nonsingular,

$$|\mathbf{A}_{11}| = \frac{\prod_{k=1}^{q} \rho_k}{[1 + \sum_{k=1}^{q} \rho_k]^m} > 0.$$

Under $\mathrm{H}_0 : \boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_q')' = \mathbf{0}$,

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}'(t)] \\ \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}(t)] & \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}(t)\boldsymbol{h}'(t)] \end{pmatrix}$$

and

$$\boldsymbol{V} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} \otimes Var[\boldsymbol{h}(t)] \end{pmatrix}$$

$$(\star) \quad \mathcal{X}_1 = n\hat{\boldsymbol{\beta}}'(\boldsymbol{A}_{11} \otimes Var[\boldsymbol{h}(t)])\hat{\boldsymbol{\beta}} \tag{3}$$

$Var[\boldsymbol{h}(t)]$ is the covariance matrix of $\boldsymbol{h}(t)$, and all moments are evaluated with respect to the reference distribution.

$$\mathcal{X}_1 \longrightarrow \chi^2_{(qp)}$$

Testing Linear Hypotheses (vector β)

$$\mathcal{X}_2 = n(\boldsymbol{H}\hat{\theta} - \mathbf{c})'(\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}')^{-1}(\boldsymbol{H}\hat{\theta} - \mathbf{c}) \quad (4)$$

where $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_q, \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_q')'$, $\boldsymbol{H}$ is $p' \times [(1+p)q)]$ predetermined matrix of rank p', $p' < (1+p)q$, $\mathbf{c}$ is a vector in $\Re^{p'}$, and the variance-covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{S}^{-1}$. It follows under $\mathrm{H}_0 : \boldsymbol{H}\boldsymbol{\theta} = \mathbf{c}$ that χ_2 is asymptotically distributed as χ^2 with (p') degrees of freedom provided the inverse exists , and H_0 is rejected for large values.

Likelihood Ratio Test ($\ell = l$)

$$
\begin{aligned}
LR &= -2[\ell(0,0) - \ell(\hat{\alpha}, \hat{\boldsymbol{\beta}})] = \\
&- 2\sum_{i=1}^{n} \log[1 + \rho_1 \hat{w}_1(t_i) + \ldots + \rho_q \hat{w}_q(t_i)] \\
&+ 2\sum_{i=1}^{q}\sum_{j=1}^{n_i}[\hat{\alpha}_i + \hat{\boldsymbol{\beta}}_i' \boldsymbol{h}(x_{ij})] + 2n\log\left[1 + \sum_{i=1}^{q}\rho_i\right]
\end{aligned}
$$

Under $\mathrm{H}_0 : \boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_\mathrm{q}')' = \mathbf{0}$, LR is asymptotically approximately distributed as $\chi^2_{(qp)}$.

Kernel density estimation (Fokianos 2004)

Consider $m = 2$. Smoothing the increments of $\hat{G}_l$ for $l = 1, 2$ and setting $w_1 \equiv \exp(\alpha_1 + \beta_1' \boldsymbol{h}(x))$, $w_2 \equiv 1$, suggests the kernel estimator,

$$\hat{g}_l(x) = \frac{1}{h_n} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \hat{p}_{ij} \hat{w}_l(x_{ij}) K\left(\frac{x - x_{ij}}{h_n}\right), \quad l = 1, 2,$$

where h_n is a sequence of bandwidths such that $h_n \to 0$ as $n \to \infty$, and K is a kernel function.

Fact: The semiparametric kernel density estimator has the same asymptotic bias but smaller variance as compared with the traditional kernel density estimator.

$$\tilde{g}_l(x) = \frac{1}{n_2 h_n} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \frac{w_l(x_{ij})}{\sum_{k=1}^{m} \rho_k w_k(x_{ij})} K\left(\frac{x - x_{ij}}{h_n}\right)$$

gives the approximation,

$$\hat{g}_l(x) = \tilde{g}_l(x) + O_p(n^{-1/2}), \quad l = 1, 2.$$

and as $n \to \infty$ and $h_n \to 0$ such that $nh_n \to \infty$,

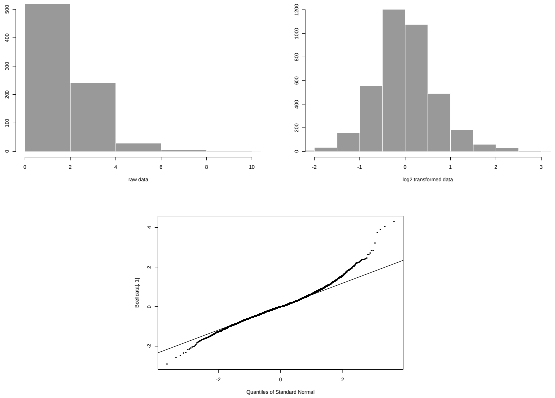
$$\mathrm{E}\left[\tilde{g}_l(x)\right] = g_l(x) + \frac{1}{2} h_n^2 g_l''(x) k_2 + o(h_n^2)$$

$$\mathrm{Var}\left[\tilde{g}_l(x)\right] = \frac{1}{n_l h_n} \frac{\rho_l w_l(x) g_l(x)}{\sum_{k=1}^{m} \rho_k w_k(x)} \int K^2(t) dt + o((nh_n)^{-1}),$$

Outline

Application to microarrays (Qi 2002)

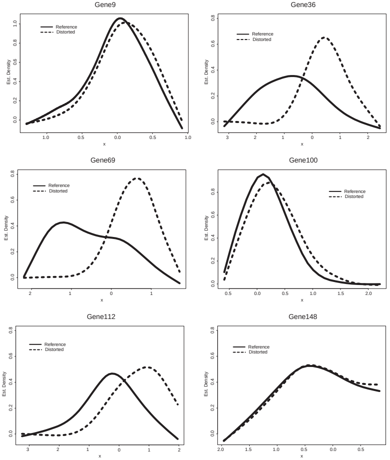
Histograms of gene expression GCAC4026 raw and $\log_2$-data. QQ-plot of the $\log_2$-data shows the data are far from normal.



Two groups, GC, AC (reference), of $\log_2$ gene expression data. $m = 2, h(x) = x$. We get practically the same results with $\boldsymbol{h}(x) = (x, x^2)'$. Plots of the estimated pdf's support the numerical results. Use $\mathcal{X}_1$, df=1, in testing similarity.

Gene	n_1	n_2	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
9	24	23	-0.049(0.06)	0.810(0.74)	1.46	0.226
36	24	21	0.452(0.35)	3.199(1.02)	97.17	0.000
69	24	22	-0.073(0.30)	2.415(0.71)	47.35	0.000
100	23	22	-0.159(0.13)	0.751(0.63)	1.24	0.265
112	24	23	-0.241(0.21)	1.698(0.53)	31.86	0.000
148	23	23	0.030(0.13)	0.100(0.44)	0.05	0.821

Spline-Smoothed Estimated Probability Densities



1

Application to radar meteorology (KWF 2004)

Reflectivity data obtained from two different radars (or "algorithms" or "sensors") at two different time periods. Data are random samples of reflectivity.

Kwajalein radar: S-band (10 cm) KPOL radar, located on Kwajalein Island at the southern end of the Kwajalein Atoll, Marshall Islands.

Brown Radar: C-band radar aboard NOAA ship Ronald H. Brown (RHB) at sea near Kwajalein Island.

The data obtained during the first period are referred to suggestively as **Kwajalein1, Brown1**, and those from the the second period are called **Kwajalein2, Brown2**.

Kwajalein1, Brown1 (Reference)

$m = 2$, $n_1 = n_2 = 500$. The hypothesis that the data come from the same radar (algorithm) is **rejected** quite conclusively.

$h(x)$	Data	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
x	1	0.784	-0.027	14.503	1.399e-03
	2	1.244	-0.042	33.476	7.216e-09
	3	0.707	-0.024	12.204	4.768e-04
	4	0.909	-0.030	17.292	3.206e-05
$\log(x)$	1	1.319	-0.396	6.520	0.011
	2	1.908	-0.575	12.744	3.572e-04
	3	1.871	-0.562	11.202	8.169e-04
	4	2.050	-0.621	16.510	4.838e-05

Brown1, Brown1 (reference)

$m = 2$, $n_1 = n_2 = 500$. The hypothesis that the data come from the same radar (algorithm) is **accepted** quite conclusively.

$h(x)$	Data	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
x	1	-0.078	0.003	0.140	0.709
	2	0.005	-0.000	0.001	0.980
	3	-0.139	0.005	0.457	0.499
	4	0.112	-0.004	0.274	0.601
$\log(x)$	1	-0.584	0.175	1.723	0.189
	2	0.095	-0.028	0.042	0.838
	3	-0.225	0.067	0.250	0.617
	4	-0.027	0.008	0.003	0.959

Kwajalein2, Brown2 (reference)

$m = 2$, $n_1 = n_2 = 500$. The hypothesis that the data come from the same radar (algorithm) is **rejected** quite conclusively.

$h(x)$	Data	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
x	1	5.323	-0.164	88.332	0
	2	3.975	-0.123	52.279	4.815e-13
	3	4.695	-0.146	74.950	0
	4	5.016	-0.156	85.325	0
$\log(x)$	1	14.359	-4.142	54.526	1.534e-13
	2	18.625	-5.367	79.723	0
	3	14.880	-4.302	60.788	6.328e-15
	4	13.580	-3.921	49.771	1.727e-12

Kwajalein2, Kwajalein2, Kwajalein2 (reference)

$m = 3$, $n_1 = n_2 = n_3 = 500$. The hypothesis that the data come from the same radar (algorithm) is **accepted** quite conclusively.

Data	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\mathcal{X}_1$	p-value
$h(x) = x$						
1	0.108	0.049	-0.003	-0.002	0.283	0.868
2	0.065	-0.003	-0.002	0.000	0.135	0.935
3	0.227	-0.041	-0.007	0.001	1.896	0.388
4	0.239	-0.220	-0.008	0.007	4.707	0.095
$h(x) = \log x$						
1	0.453	2.278	-0.132	-0.665	1.929	0.381
2	-0.792	-0.223	0.231	0.065	0.250	0.882
3	-0.359	0.735	0.105	-0.215	0.553	0.758
4	1.665	1.246	-0.485	-0.363	1.014	0.602

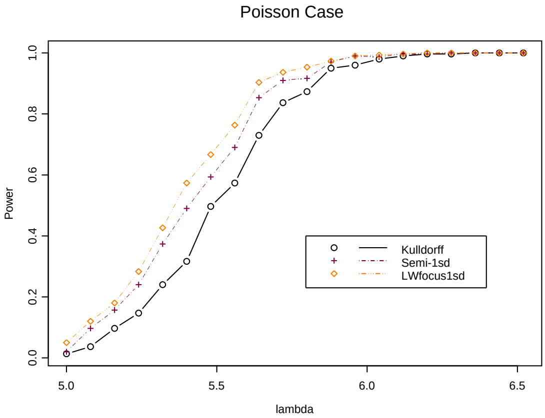
Application to cluster detection (KW 2007)

(•)A cluster is a subregion where the distribution of the variable of interest is different from the distribution in the rest of the area.

(•)Assume the location and the size of the cluster are known. We test whether the distributions inside and outside the cluster are the same.

(•)Power comparison vs Kulldorff (1997,2001), Lawson-Waller (1995) focused test.

Specialized case Poisson: Power curves for one-sided tests in the Poisson case. Scalar β, $h(x) = x$.

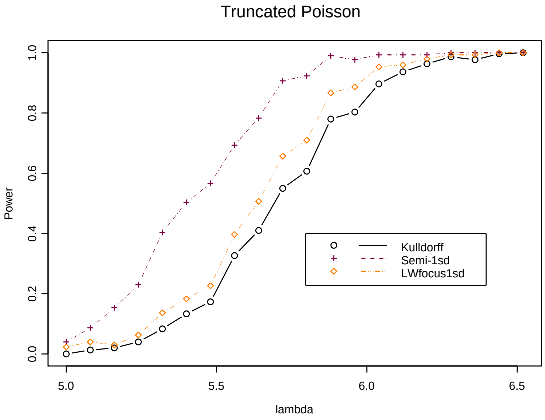


Clipped Poisson: $x \sim$ Poisson(λ).

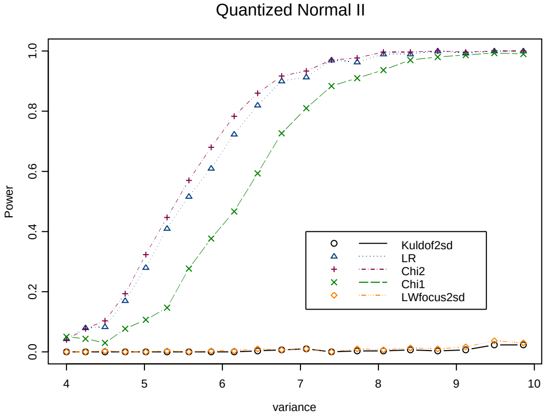
$$z = \begin{cases} 2 & (x <= 2) \\ x & (2 < x <= 10) \\ 10 & (x > 10) \end{cases}$$

Kulldorff's and the focused tests applied under Poisson.

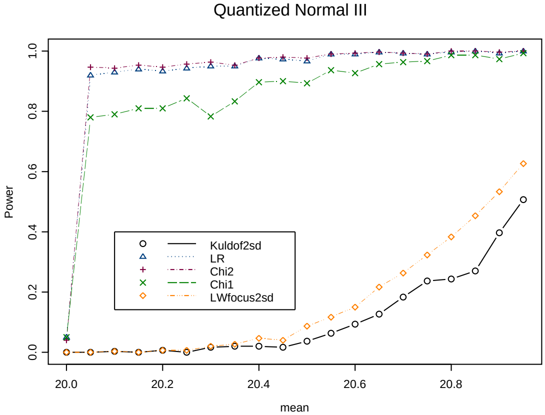
Power curves for one-sided tests in a clipped Poisson case. Scalar β, $h(x) = x$.



Power curves for two-sided tests applied to quantized normal data with the different means and different variances. $m = 2, \boldsymbol{h}(x) = (x, x^2)'$.



Power curves for two-sided tests applied to quantized normal data with the same means but different variances.
$m = 2, \boldsymbol{h}(x) = (x, x^2)'$.



Semiparametric time series prediction

Assume a system with m regressions:

$$
\begin{aligned}
x_{1t} &= \hat{f}_1(\mathbf{z}_{1,t-1}) + \epsilon_{1t}, \quad t = 1, ..., n_1 \\
&\cdot \\
&\cdot \\
&\cdot \\
x_{qt} &= \hat{f}_q(\mathbf{z}_{q,t-1}) + \epsilon_{qt}, \quad t = 1, ..., n_q \\
x_{mt} &= \hat{f}_m(\mathbf{z}_{m,t-1}) + \epsilon_{mt}, \quad t = 1, ..., n_m
\end{aligned}
$$

- The x_{jt} may be nonstationary, and the f_i may be nonlinear.
- The ϵ_{kt} are the residuals. They may be dependent.

Suppose that for each t,

$$\epsilon_{jt} \sim g_j(x), \quad j = 1, ..., q, m$$

Define the *reference* density: $g(x) = g_m(x)$
We assume:

$$g_j(x) = \exp\{\alpha_j + \boldsymbol{\beta}_j' \mathbf{h}(x)\} g(x), \quad j = 1, ..., q \tag{5}$$

Define *the combined residual data*

$$\boldsymbol{\tau} \equiv \left\{ (\epsilon_{1t}, ..., \epsilon_{1n_1}), ..., (\epsilon_{q1}, ..., \epsilon_{qn_q}), (\epsilon_{m1}, ..., \epsilon_{mn_m}) \right\}$$

Problem: From *the fused residual data* estimate all the $\alpha_j, \boldsymbol{\beta}_j$, and $g(x)$,$G(x)$, for the purpose of predicting the future reference value $x_{m,t+1}$.

Prediction

Since

$$x_{m,t+1} = \hat{f}_m(\boldsymbol{z}_{m,t}) + \epsilon_{m,t+1}$$

and $\epsilon_{m,t+1} \sim G$, we have the useful approximation of the predictive probability at $t+1$ conditional on $\boldsymbol{z}_{m,t}$,

$$\begin{aligned} P(x_{m,t+1} \leq x \mid \boldsymbol{z}_{m,t}) &= G(x - \hat{f}_m(\boldsymbol{z}_{m,t})) \\ &\approx \hat{G}(x - \hat{f}_m(\boldsymbol{z}_{m,t})) \end{aligned} \tag{6}$$

From (6) we can get the predicted values

$$\hat{E}(x_{m,t+1} \mid \boldsymbol{z}_{m,t})$$

SP Prediction in Bivariate AR

$$x_t = a_1 x_{t-1} + a_2 y_{t-1} + \epsilon_t$$
$$y_t = b_1 x_{t-1} + b_2 y_{t-1} + \eta_t$$

Independent Gaussian noise components $\epsilon_t \sim \mathrm{N}(0, \sigma_1^2)$ and $\eta_t \sim \mathrm{N}(0, \sigma_2^2)$. We have:

$$\begin{aligned} g_\epsilon(x) &= \exp\left\{\log \frac{\sigma_2}{\sigma_1} + \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right) x^2\right\} g_\eta(x) \\ &\equiv e^{\alpha+\beta x^2} g_\eta(x) \end{aligned}$$

with $m = 2, q = 1$.

The bivariate AR(1) system (7) was simulated with parameters

$$(a_1, a_2, b_1, b_2, \sigma_1, \sigma_2) = (0.6, -0.5, 0.4, 0.5, 1, 0.5)$$

and $n_1 = n_2 = 500$.

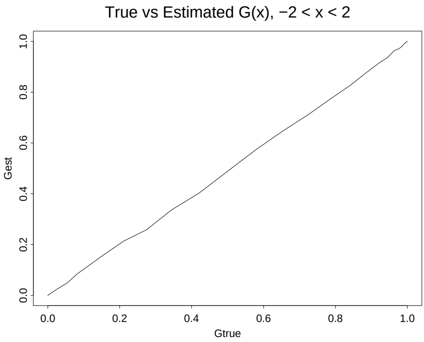
The true “tilt” parameters are:

$$(\alpha_1, \beta_1) = (-0.693, 1.5)$$

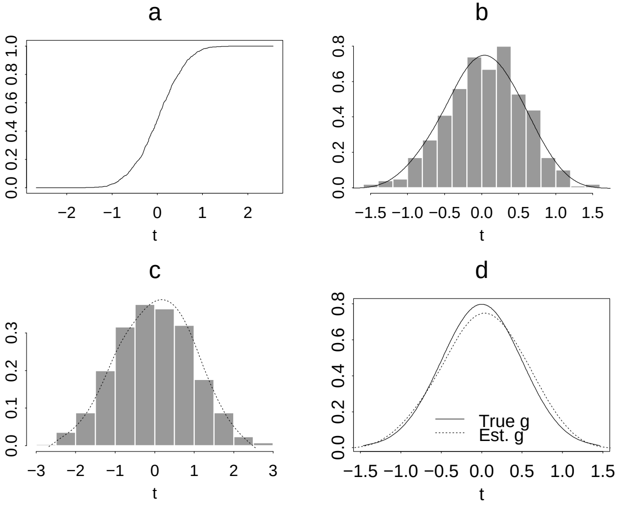
and the reference g is the pdf of η_t,

$$\eta_t \sim \mathrm{N}(0, 0.5^2)$$

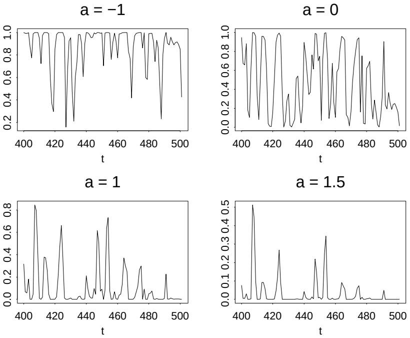
Pairs of $(G(x), \hat{G}(x))$, $-2 < x < 2$.



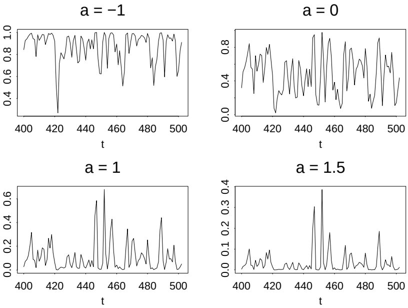
a. Ref. $\hat{G}_\eta$ b. $\hat{g}_\eta$ and hist. of η_t. c. Tilted pdf $\hat{g}_\epsilon$ and hist. of ϵ_t. d. Est. and ref. $g = g_\eta$.



One step ahead prediction of threshold probabilities:
$\hat{P}(y_t > a \mid x_{t-1}, y_{t-1})$, $t = 400, ..., 501$.



Two steps ahead prediction of threshold probabilities: $\hat{P}(y_t > a \mid x_{t-2}, y_{t-2})$, $t = 400, ..., 502$.



LA mortality prediction

Filtered mortality data from Los Angeles County, 01.01.1970 to 12.31.1979, discussed in Shumway et al(1988). The original daily data, consisting of the response series (total mortality) and its covariate series (two weather and six pollution series), were filtered and then sampled weekly to produce series of length $N = 508$ each.

$y = \text{Total mortality}$
$T = \text{Temperature (centered at 74.26 F}^\circ)$
$CO = \text{Carbon monoxide}$

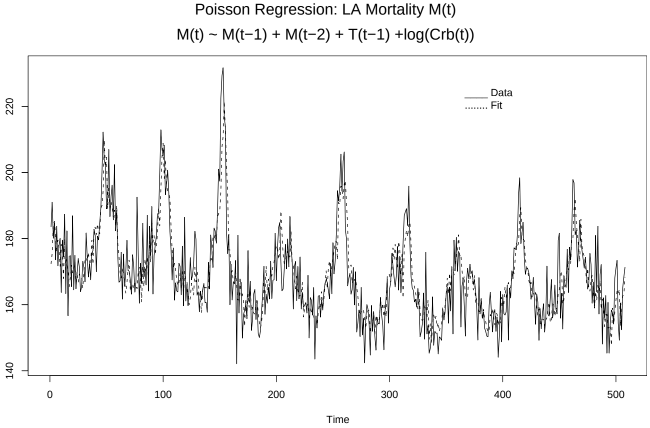
Model 1 (Kedem and Fokianos (2002)).

$$y_t = \exp\left\{\hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \hat{\beta}_3 T_{t-1} + \hat{\beta}_4 \log(CO_t)\right\} + \eta_t$$

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4) = (4.505, 0.00189, 0.00184, -0.00133, 0.04683)$$

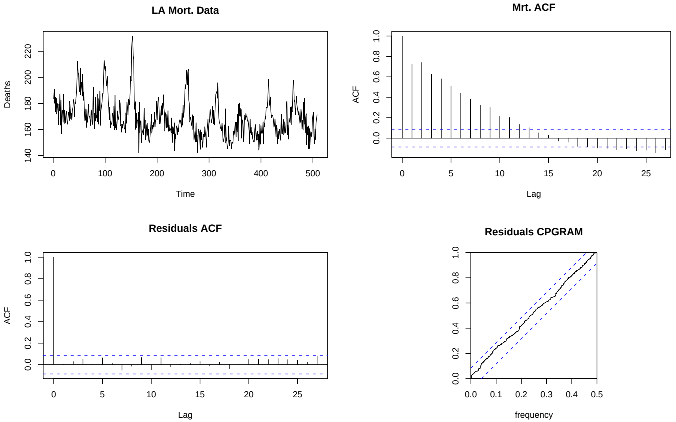
The exponential model outperforms several competitors, giving close to white noise residual η_t.

LA Mortality Data vs Fit.



LA Mortality residuals:

Poison Reg: M(t) ~ M(t−1) + M(t−2) + T(t−1) +log(Crb(t))



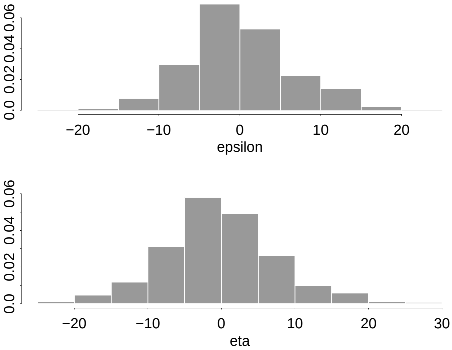
Additional regression: **Model 2**

$$T_t = \hat{\phi}_1 T_{t-1} + \hat{\phi}_2 T_{t-2} + \hat{\phi}_3 T_{t-3} + \hat{\phi}_4 T_{t-4} + \epsilon_t$$

$$(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4) = (0.272, 0.283, 0.096, 0.183)$$

The sample autocorrelation of the residual ϵ_t has somewhat significant values at lags 8,9,22, so that ϵ_t is probably not white noise.

LA Mortality: Histograms of ϵ_t and η_t residual series.

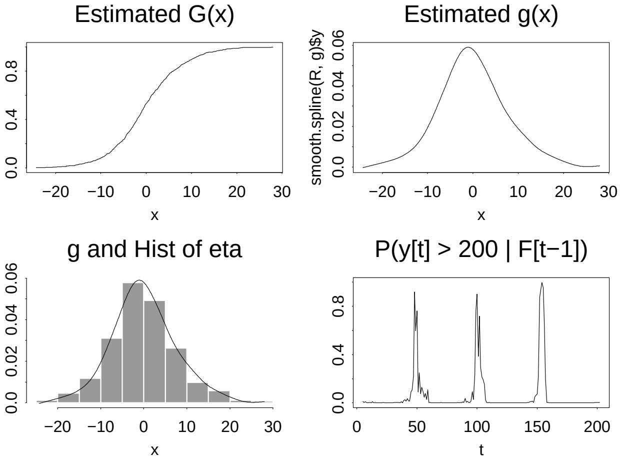


The histograms suggest the tilt model

$$g_\epsilon(x) = e^{\alpha+\beta x^2} g_\eta(x)$$

where η_t is the reference with density $g = g_\eta$ and distribution function G. In the present case

$$(\hat{\alpha}, \hat{\beta}) = (0.145(0.04), -0.003(0.00086))$$


Estimated G(x)
0.0
0.4
0.8
−20
−10
0
10
20
30
x
Estimated g(x)
smooth.spline(R, g)$y
0.00
0.02
0.04
0.06
−20
−10
0
10
20
30
x
g and Hist of eta
0.00
0.02
0.04
0.06
−20
−10
0
10
20
30
x
P(y[t] > 200 | F[t−1])
0.0
0.4
0.8
0
50
100
150
200
t

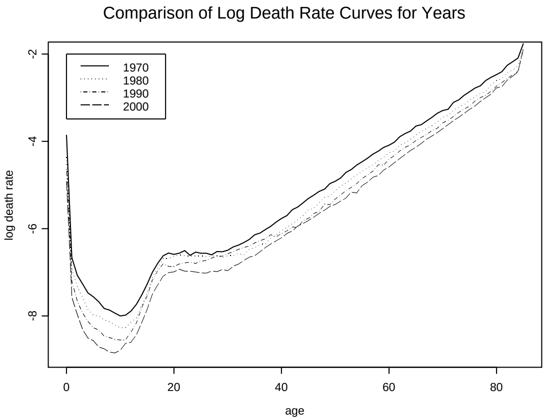
Bottom right: The estimated prediction probabilities obtained from $\hat{G}$ of exceeding 200 “deaths”,

$$P(Y_t > 200 \mid Y_{t-1}, Y_{t-2}, T_{t-1}, \log(CO_t)) \approx 1 -$$
$$\hat{G}\left(200 - \exp\left\{\hat{\beta}_0 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 Y_{t-2} + \hat{\beta}_3 T_{t-1} + \hat{\beta}_4 \log(CO_t)\right\}\right)$$

t=3,...,202, pointing to a strong annual cycle.

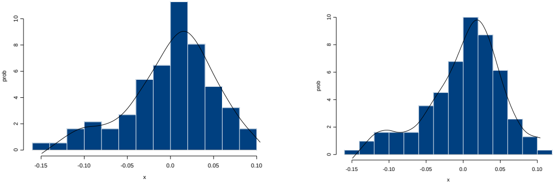
Forecasting U.S. mortality (KLWW 2008)

Mortality curves: Log death-rate as a function of age.

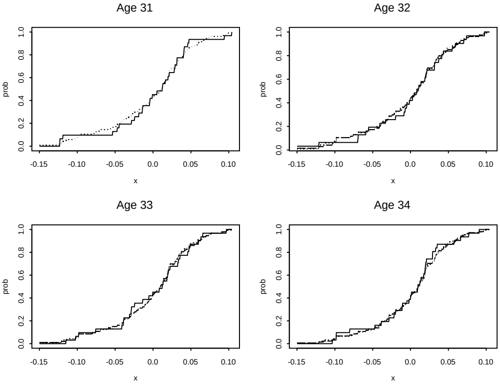


- For each age 1-85 we have a short ts: 1970-2001 (size 32).
- Fusion of short time series 1970-2001 and predicting 2002.
- Each age is predicted from 5 adjacent time series corresponding to 5 ages.
- Autoregression for each age: $x_t = bx_{t-1} + c + \epsilon_t$.

Estimated reference pdf of age 33 from the combined residual data for the 3-age group 32-34 (left), and the 5-age group 31-35 (right). Since we combined more information in the 5-age group there is a noticeable improvement in the density fit.



5-age group 31-35: Comparison of the empirical (solid line, 32 resid.) and estimated (dotted line, 5*32=160 resid.) residual cdf's for the indicated ages. The estimated cdf for age 35 is not shown.



All Age MSE: Lee & Carter (1992) SVD vs SP.

Prediction	Case	LC	SP
1 Step	Total Pop	0.297	0.104
1 Step	Female	0.619	0.187
1 Step	White Female	0.645	0.249
2 Step	Total Pop	0.389	0.180

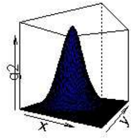
Case-control in testicular germ cell data (KKVG 2009)

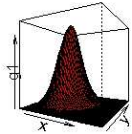
Bivariate (Height,Weight) data: $\boldsymbol{x} = (x, y)'$.
Assume the tilt model

$$g_1(\boldsymbol{x}) = \exp(\alpha_1 + \beta_1' \boldsymbol{x}) g_2(\boldsymbol{x})$$

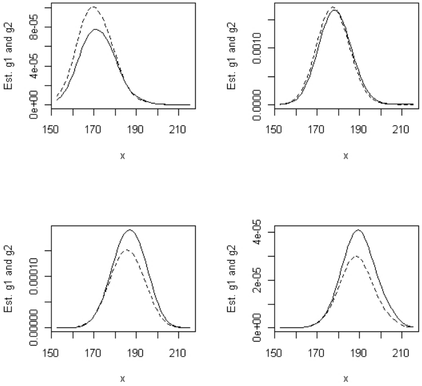
- $LR = 15.108$, $df = 2$, $p - value = 0.0005$.
- Thus, when height and weight are considered jointly, we reject the null hypothesis $H_0 : \beta_1 = 0$ of equidistribution quite conclusively, so that height and weight are significant risk factors.

Estimated densities of case(g_1) and control(g_2).





Height distribution: Overlay of vertical slice cuts. Weight (y value) was fixed at 52.555 (top left), 77.555 (top right), 107.555 (bottom left), and 117.555 (bottom right) kg.



Weight distribution: Overlay of vertical slice cuts. Height (x value) was fixed at 161.4(top left), 171.4(top right), 191.4(bottom left), and 196.4(bottom right) cm.

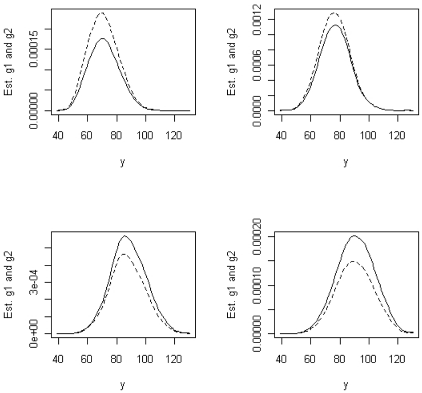


Table: Some joint probabilities of height and weight in the case and control groups.

Probability	Case	Control
Pr(H $\leq$ 155, W $\leq$ 59)	0.000769	0.000374
Pr(H $\leq$ 165, W $\leq$ 59)	0.005750	0.003490
Pr(H $\leq$ 178, W $\leq$ 65)	0.066406	0.051604
Pr(H $\leq$ 185, W $\leq$ 70)	0.161664	0.133651
Pr(H $\leq$ 180, W $\leq$ 80)	0.375041	0.315808
Pr(H $\leq$ 180, W $\leq$ 90)	0.520636	0.448033
Pr(H $\leq$ 187, W $\leq$ 95)	0.818147	0.769016
Pr(H $\leq$ 200, W $\leq$ 100)	0.957770	0.943891
Pr(H $\leq$ 203, W $\leq$ 119)	0.996346	0.993959

Some Open Research Problems

1. Multisensor Data Fusion.
2. A general approach to regression: Estimate the conditional expectation of the response given the covariates. Diagnostics.
3. A general approach to classification: Get rid of the normal assumption.
4. A general approach to ANOVA.
5. Incorporate state-space models.
6. General Mixed Models.