Robust Axial Data Analysis

Daijin Ko and Ted Chang
University of Texas at San Antonio
and University of Virginia

Robustness

- Robustness
- SB Robustness on spheres
- Robustness on Real projective space
- M-estimation of axial location

Robust Statistics

- The question: are statistical methods, which are good under the model, reasonably good even if the data is contaminated?
- Does contaminated data have a big effect on the studied estimator?
- What effect?

What effect?

- The bias of a statistic
- Variance of a statistic (efficiency)
- One can control the variance by increasing sample size but cannot control bias

Bias Control

- The **influence function** of statistical functional T is defined as

$$IF(x, T, F) = \lim_{s \to 0} \frac{(T((1 - s)F + s\delta_x) - T(F))}{s}$$

where δ_x denotes the point mass at x

- The **gross error sensitivity** of T at F is defined by the supremum of the norm of IF taken over all x

$$GES(T, F) = sup_x |IF(x, T, F)|$$

measure of robustness

- The gross error sensitivity can be regarded as an approximate upper bound for the asymptotic bias of an estimator T

- The GES measures the largest influence that a small amount of contamination can have (to the bias)

- If the gross error sensitivity of a functional is finite, then, the "corresponding" estimator is called B-robust (for Bias)

Spherical mean

- The spherical mean T

$T(F) = (\int x dF_1, \ldots, \int x dF_q) / \| (\int x dF_1, \ldots, \int x dF_q) \|$

- It is the MLE of the FVML distribution

- $IF(z; T, F) =$

$(q-1)\left(z - (z^t T(F))T(F)\right) / E\, g'(t)(1-t^2)$

where $t = \mu^t X$, X is a random unit vector with distribution *F* (*density f*) and *g(s)=log(f(s)).*

The GES measures at FVML

When F = *FVML* (μ, κ),

$\| IF(z;T,F) \|^2 = (q-1)^2(1-s^2)/E^2 g'(t)(1-t^2)$

where $s = z^t\mu$ and $E g'(t)(1-t^2) = (q-1)\, A_q(\kappa)/\kappa$

$GES(T,F) = \sup_Z \| IF(z;T,F) \| = 1/A_q(\kappa)$, *bounded*.

- *Is the mean B-robust?*

When $\kappa > 0$, the spherical mean is robust??

But of course, on a bounded parameter space, this bias is bounded.

SGES

- The idea is that on such bounded parameter space, we should rather compute the influence of a contamination in the unit of a scale functional S of the distribution
- *Standardized influence function*

$$SIF(z;T,S,F) = \frac{1}{S(F)} IF(z;T,F)$$

- The standardized gross error sensitivity at a family $\mathcal{F}$ of distributions is defined

$$\text{SGES}(T, S, \mathcal{F}) = \sup_{F \in \mathcal{F}} \text{GES}(T, F) / S(F)$$

SB Robustness

- The standardized gross error sensitivity at a family $\mathcal{F}$ of distributions is defined

$$\text{SGES}(T, S, \mathcal{F}) = \sup_{F \in \mathcal{F}} \text{GES}(T, F) / S(F)$$
$$= \sup_{F \in \mathcal{F}} \sup_{Z} \| IF(z; T, F) \| / S(F)$$

- It measures the maximum asymptotic bias of T in the units S(F) within the family $\mathcal{F}$

- T is called SB-robust at $\mathcal{F}$ if SGES(T, S, $\mathcal{F}$) is bounded

SGES

- At $\mathcal{F}$={F}, SB-robustnes is B-robustness

- When $\mathcal{F}$ ={*FVML* (μ, κ) | κ>0}, for S^{-2} (F) is the Fisher Information for location of F,

$$\text{SGES} = \sup_{\kappa>0} (1/A_q(\kappa))\sqrt{\kappa A_q(\kappa)}$$
$$= \sup_{\kappa>0} \sqrt{\kappa/A_q(\kappa)} = \infty$$

- Therefore, the asymptotic bias of T associated with a small contamination could be very large compare to the dispersion.

Scale

The standardized gross error sensitivity at a family $\mathcal{F}$ of distributions is defined in general

SGES(T, $\mathcal{F}$) =

$$\sup_{F \in \mathcal{F}} \sup_{x} \left(IF(x,T,F)^t S^{-1}(F) IF(x,T,F) \right)$$

where S^{-1} (F) is the Fisher Information for T(F)

Fisher Information metric

Let v and w be tangent vectors to S^{k-1} at some point $\theta \in S^{k-1}$ such that $v = \gamma_1'(0)$ *and* $w = \gamma_2'(0)$ for some curves γ_1, and γ_2 are curves on S^{k-1} such that $\theta = \gamma_1(0) = \gamma_2(0)$.

Let $f(\theta' x)$ be the density of X. Define

$$<v,w>_\theta = \frac{\partial^2}{\partial s \partial t}\bigg|_{(t,s)=(0,0)} E[\log(f(\gamma_1(s)^t X)\log(f(\gamma_2(t)^t X))]$$

Then $<v,w>_\theta = c_f v^t w = (E[\varphi_f(\theta' X)(1-(\theta' X)^2)]/(k-1) v^t w$

i.e., $c_f = (E[\varphi_f(\theta' X)(1-(\theta' X)^2)]/(k-1)$.

Fisher Information metric

Since IF (x,T,F) is tangent vector to S^{k-1} at $T(F) \in S^{k-1}$

$$SGES = \sup_F \sup_x < \mathrm{IF}\,(x,T,F),\, \mathrm{IF}\,(x,T,F) >_{T(F)}$$

Fisher Information metric

For rotationally symmetric case,

$$\mathrm{SGES}(\mathrm{T}, F) = \sup_F \sup_x c_f \left\| IF(x, T, F) \right\|.$$

The definitions of SGES are equivalent for ($\mathrm{S(F)} = c_f^{-1/2}$).

SB robust M-estimator on spheres

- The M-estimators proposed by Ko and Chang (93) are then defined by

$$\hat{\theta} = argmin_{\theta \epsilon \Omega_q} \, n^{-1} \sum_{i=1}^{n} \rho(X_i, \theta)$$

- This is a constrained maximization problem so using the Lagrange multipliers method we obtain the estimating equation

$$\mathrm{M}(\mathrm{s}, \phi, F_n) = E_{F_n}\phi(\mathrm{X}, \mathrm{s}) - |E_{F_n}\phi(\mathrm{X}, \mathrm{s})|\,\mathrm{s} = 0$$

where $\phi(\mathrm{x}, \theta) = (\partial/\partial\theta)\, \rho(\mathrm{x}, \theta)$.

Influence function of M-estimator

- M(s, ψ, F) = E_F [ψ(T(F)′X)(X – (X′T(F))T(F))] = 0

- $IF(z;T, F) = \dfrac{(q-1)\psi(\theta^t z)(z-(\theta^t z)\theta)}{E\psi(t)g'(t)(1-t^2)}$,

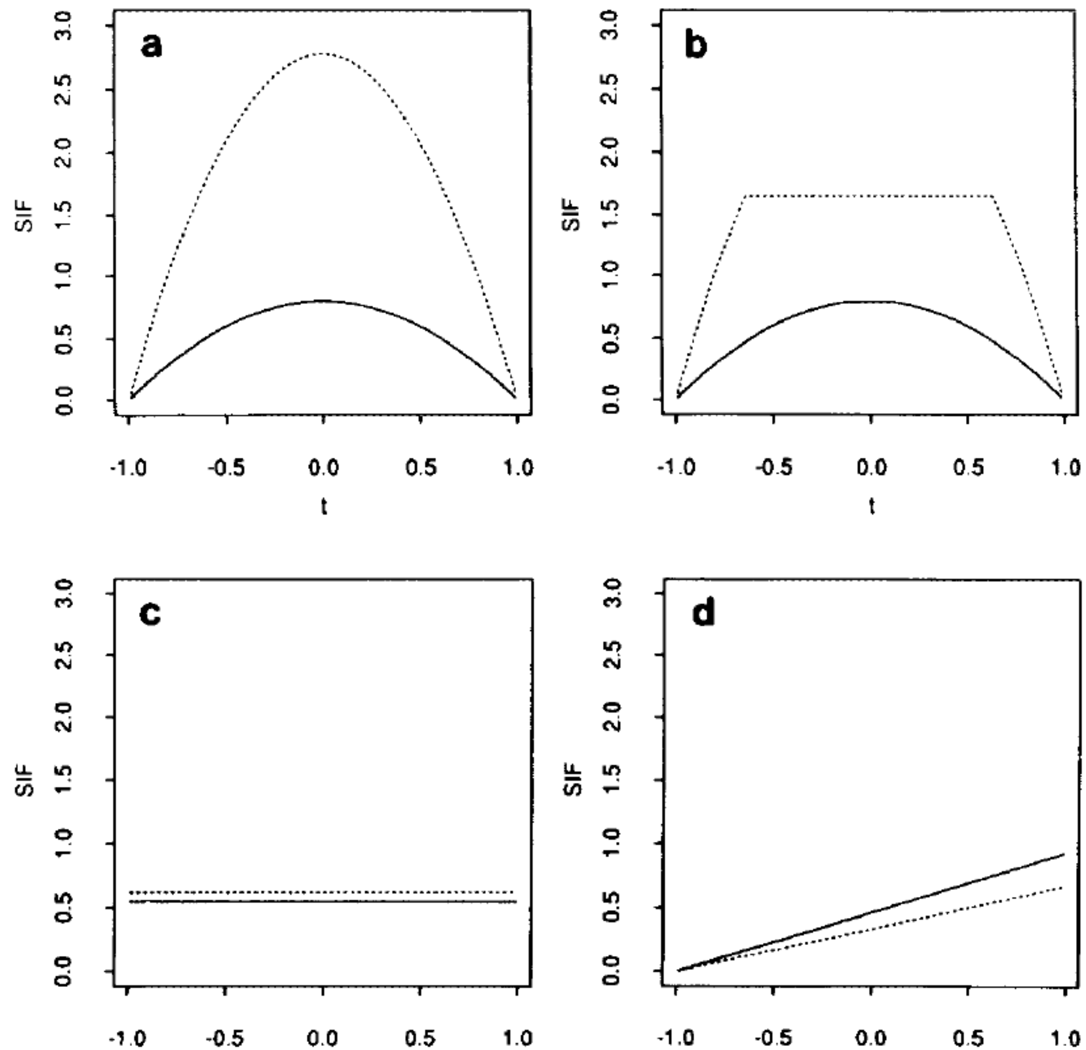
  where t = $\theta^t X$ and X is a random unit vector with rotationally symmetric distribution F around θ.

SGES at VMFL(μ, κ)

- SIF = $\frac{(q-1)\psi(s)\sqrt{1-s^2}}{\kappa E\psi(t)(1-t^2)}\sqrt{\kappa A_q(\kappa)}$
- $SGES = sup_{-1\leq s\leq 1}\frac{(q-1)\psi(s)\sqrt{1-s^2}}{\kappa E\psi(t)(1-t^2)}\sqrt{\kappa A_q(\kappa)}$

SIF for κ=1 (dotted) and κ=10 (solid)

a) **Spherical Mean**
b) **Optimal M-estimator with 90% efficiency**
c) **Spherical Median**
d) **Normalize spatial median**



SB-robustness of M estimators

Spherical Mean: Not SB-robust

Optimal M-estimator: SB-robust if b=O$(\kappa^{-1/2})$.

Spherical Median: SB-robust

Normalize spatial median: SB-robust

Asymptotic Distribution

$\hat{\mu}_n$*: an M-estimator, the solution of the M Equation*
$\sum_{i=1}^{n} \psi(\hat{t}_i)(X_i - \hat{t}_i\mu) = 0$

- Under regularity conditions, M-estimators are consistent and asymptotically normal.

- $\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} N(0, {c_1}^{-2}\Sigma)$ where

$\Sigma = E(\psi(t)^2(1 - t^2))(I - \mu\mu^t)/(q-1)$ and

$c_1 = E[\psi(t)g'(\mathrm{t})(1 - t^2)]/(q-1)$.

Asymptotic Distribution

n$(\hat{\mu}_n$-$\mu)^2 \overset{d}{\rightarrow}$

$E(\psi(t)^2(1-t^2))$/ $E[\psi(t)g'$(t)$(1-t^2)]$*$\chi^2{}_{q-1}$

One can construct the (1-α)-confidence cone for the true parameter μ.

Axial Data

$\boldsymbol{R}P^{p-1}$: real projective p-1 dim space

[x]={x, -x} $\epsilon \boldsymbol{R}P^{p-1}$ = ${S^{p-1}}_{x\sim -x}$

$\boldsymbol{R}P^{p-1}$ is $S^{p-1}$with x identified with –x using the the "quotient topology".

ϕ: $S^{p-1} \rightarrow \boldsymbol{R}P^{p-1}$, natural projection

ϕ(x)= [x]={x, -x}

Axial Data

ϕ: $S^{p-1} \rightarrow \boldsymbol{R}P^{p-1}$, natural projection

ϕ(x)= [x]={x, -x}

ϕ: double covering

i.e., there is a 1-1 correspondence between continuous, respectively differentiable, maps f on S^{p-1} that satisfy the condition f(x)=f(-x) and continuous , respectively differentiable, maps $\tilde{f}$ on $\boldsymbol{R}P^{p-1}$ so that f= $\tilde{f}$ ϕ.

Tangent Spaces

$T_{\theta_0}S^{p-1}$= {x $\in R^{p-1}$ | x$^t\theta_0$= 0}

: the tangent space to S^{p-1} at $\theta_0 \in S^{p-1}$.

$T_{\theta_{0,}-\theta_0}\boldsymbol{R}P^{p-1}$, the tangent space to $\boldsymbol{R}P^{p-1}$ at $\{\theta_0, -\theta_0\} \in \boldsymbol{R}P^{p-1}$

The map ϕ defines a vector space isomorphism $\phi_*: T_{\theta_0}S^{p-1} \rightarrow T_{\theta_{0,}-\theta_0}\boldsymbol{R}P^{p-1}$

Riemannian metric on $\boldsymbol{R}P^{p-1}$

For $\tilde{v}, \tilde{w} \epsilon T_{\theta_{0,}-\theta_0} \boldsymbol{R}P^{p-1}$, let v, w $\in T_{\theta_0} S^{p-1}$ be such that $\phi_*(v)$= $\tilde{v}$ and $\phi_*(w)$= $\tilde{w}$.

Define a Riemannian metric <,> $_0$ on $\boldsymbol{R}P^{p-1}$ by < $\tilde{v}, \tilde{w}$ > $_0$= $v^t\, w$.

Riemannian metric is a positive definite inner product on each tangent space subject to a differentiability constraint.

Riemannian metric on $\boldsymbol{R}P^{p-1}$

Using this definition, each $\phi_*: T_{\theta_0}S^{p-1} \to T_{\theta_{0,}-\theta_0}\boldsymbol{R}P^{p-1}$ is an isometry.

Riemannian metrics define a volume element, and under this choice of Riemannian metric for $\boldsymbol{R}P^{p-1}$, the (surface) volume of each $\boldsymbol{R}P^{p-1}$ is half that ofS^{p-1}.

In particular, the surface volume of $\boldsymbol{R}P^2$ is 2π and that ofS^2 is 4π.

More generally, let $S_+^{p-1} = \{x = [x_1, x_2, \ldots, x_p] \in S^{p-1} \mid x_1 \geq 0\}$. Except on the 'equatorial S^{p-2}', ϕ is 1-1 onto RP^{p-1}. It follows from the change of variables theorem, that if $\tilde{f} : RP^{p-1} \to R^1$ then

$$\int_{RP^{p-1}} \tilde{f} = \int_{S_+^{p-1}} \tilde{f}\phi = \frac{1}{2} \int_{S^{p-1}} \tilde{f}\phi \qquad (1)$$

Riemannian metric on $\boldsymbol{R}P^{p-1}$

The relevant group is *G=SO(p)* with action

A· $\{\theta, -\theta\}$ = $\{A\theta, -A\theta\}$ where θ represent generic element of S^{p-1}. For specific A, let f:$S^{p-1} \to \boldsymbol{R}P^{p-1}$ be the map f(x)=$\{A$x$, -A$x$\}$.

This induced a map $\tilde{f}$: $\boldsymbol{R}P^{p-1} \to \boldsymbol{R}P^{p-1}$ such that f = $\tilde{f}$ ϕ and G-action *A*· $\{\theta, -\theta\}$ = $\{A\theta, -A\theta\}$ is well defined.

Let $\theta_0 \in S^{p-1}$be θ_0=[1,0,…,0] and $\tilde{\theta}_0 \in \boldsymbol{R}P^{p-1}$ be $\tilde{\theta}_0$= $\phi(\theta_0)$= {θ_0, - θ_0}.

If A· $\tilde{\theta}_0$=$\tilde{\theta}_0$, then either Aθ_0=θ_0 or Aθ_0= -θ_0 .

In the former case A has the form

$\begin{bmatrix} 1 & 0 \\ 0 & A_1 \end{bmatrix}$ (2) where $A_1 \in$ SO(p-1) and

in the latter case,

$\begin{bmatrix} -1 & 0 \\ 0 & A_1 \end{bmatrix}$ (3) where $A_1 \in$ SO_(p-1).

where SO_(p-1) = {A_1= (p-1)x(p-1) matrix | ${A_1}^t A_1$= I & det(A)=-1}

- Thus the isotropy group H is the subgroup of all matrices of the forms (2) or (3).

ϕ_* is an isomorphism $T_{\theta_0}S^{p-1} \to T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$ and $T_{-\theta_0}S^{p-1} \to T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$

We represent $T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$ as $T_{\theta_0}S^{p-1}$= R^{p-1}= {[0, e1] | e1 ∈ R^{p-1}} using ϕ_*.

Then A∈ H, A· $[0, e1] = [0, A_1$e1] if A has form (2) and A· $[0, e1] = [0, -A_1$e1] if A has form (3).

Since the representation of SO(p − 1) on R^{p-1} is irreducible the representation of H on $T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$ is *a fortiori* irreducible

- Thus in Propositions 1 and 2 of Chang and Rivest (2001), there is only one irreducible subspace and it only remains to identify the constants $c = c1$ and $d = d1$ of Chang-Rivest equation (16).

Axial Data

Consider the distribution f on S^{p-1} with density of the form f(x; $\theta) = f_0[(x^t\theta)^2]$ for x, $\theta \epsilon S^{p-1}$.

Using (1), if $\tilde{\text{f}}$ is defined by

$\tilde{\text{f}}(\phi$(x); $\phi(\theta)$) = $2f_0[(x^t\theta)^2]$, $\tilde{\text{f}}$ is a density on $\boldsymbol{R}P^{p-1}$.

Consider objective functions

$\tilde{\rho}$ (ϕ(x); $\phi(\theta)$) = $\rho_0[(x^t\theta)^2]$

M-estimator

For objective functions $\tilde{\rho}$ (ϕ(x); $\phi(\theta)$) = $\rho_0[(x^t\theta)^2]$,

The M-estimators are then defined by

$$\hat{\theta} = argmin_{\theta \epsilon S^{p-1}} \, n^{-1} \sum_{i=1}^{n} \tilde{\rho} \, (\, \phi(x_i); \; \phi(\theta))$$
$$= argmin_{\theta \epsilon S^{p-1}} \, n^{-1} \sum_{i=1}^{n} \rho_0[(x_i{}^t \theta)^2]$$

SB robust M-estimator

- This is a constrained maximization problem so using the Lagrange multipliers method we obtain the estimating equation

$$M(s, \phi, F_n) = E_{F_n}\phi(X, s) - |E_{F_n}\phi(X, s)|s = 0$$

where $\phi(x, \theta)=(\partial/\partial\theta)\ \rho(x, \theta)$.

When $\rho(x, \theta) = \rho(x^t\theta) = \tilde{\rho}\,[(x^t\theta)^2]$,

$\psi(t)$=- $\rho'(t)$ =$\tilde{\rho}'(t^2)$*2t

Influence function of M-estimator

- M(s, ψ, F) = E_F [ψ(T(F)'X)(X – (X'T(F))T(F))] = 0

- $IF(z;T, F) = \dfrac{(q-1)\psi(\theta^t z)(z-(\theta^t z)\theta)}{E\psi(t)g'(t)(1-t^2)}$,

  as before.

Robustness of M-estimator

- For MLE of Scheidegger-Watson distribution F and X~F , and ψ(t)=2t

$$IF(z;T,F) = (q-1)(q\lambda_1-1)^{-1} t\sqrt{1-t^2}$$

where λ_1 is the largest eigenvalue of EXX^t.

AT SW= $\{f(\theta,\kappa) = (w_q b_q(\kappa))^{-1} \exp\{\kappa\,(\theta^t x)^2\}\ |\kappa > 0\}$,
SGES(T, SW)=$\sup_{\kappa>0}\sup_t |IF(z;T,F|\,/O(\kappa^{-1/2})$ =∞ and MLE is not SB-robust.

SB Robustness of M-estimators

Can show that at SW = $\{f(\theta, \kappa) = (w_q b_q(\kappa))^{-1}$ exp$\{\kappa\ (\theta^t x)^2\}$ $|\kappa > 0\}$, spherical median axis, normalized spatial median, and optimal M-estimator with ψ bounded by O($k^{-1/2}$) are SB-robust.

Asymptotic Distribution of M estimators (Brown 85)

$X : sample\ space$

$\Theta : parameter\ space = R^q$

$f(x,\theta) : family\ of\ densities$

$\rho(x,\theta) : objective\ function$

$\mathrm{X} = (X_1,...,X_n)\ sample\ from\ f(x,\theta_0)$

$$\hat{\theta} = \arg\min_\theta \sum_i \rho(X_i,\theta)$$

$$S(\mathrm{X},\theta) = \sum_i \partial\rho(X_i,\theta)/\partial\theta = 0 : estimating\ equation$$

$$\nu_{\theta_0}(\theta) = E_\theta S(\mathrm{X},\theta_0)$$

$$\nu'_{\theta_0}(\theta_0) = \frac{\partial}{\partial\theta} |_{\theta=\theta_0} \nu_{\theta_0}(\theta)$$

Asymptotic Distribution of M estimators (Brown 85)

$$Under\ regularity\ conditions,$$

$$n^{1/2}(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,B^{-1}A(B^{-1})^t)$$

$$where\ A=\lim_n Cov_{\theta_0}(S(X,\theta_0))/n\ \ and$$

$$B=\lim_n \nu'_{\theta_0}(\theta_0)/n$$

Asymptotic Distribution of M estimators on Manifolds

Chang and Tsai ('99) Reformulate A and B in a coordinate-free manner for a differentiable manifold Θ.

$\gamma = (\gamma_1, \ldots, \gamma_q) : R^1 \to R^q$ *a curve*

$f : R^q \to R^1$

$$(f \circ \gamma)'(0) = \sum_i \frac{\partial f}{\partial x_i}(\gamma(0)) \cdot \gamma_i'(0)$$

depends only upon a base point $\gamma(0)$ and a tangent vector $\gamma'(0)$.

Asymptotic Distribution of M estimators on Manifolds

tan gent vector to Θ at $\theta_0 \in \Theta$ is an equivalence class of curves satisfying $\gamma(0) = \theta_0$, where two curves γ_1 *and* γ_1 are equivalent (defines the same tangent vector at θ) if $(f \circ \gamma_1)'(0) = (f \circ \gamma_2)'(0)$ for any $f : \Theta \to R^1$.

$T_{\theta_0}\Theta = \{\text{tangent vectors at } \theta_0\}$ is a vector space

Reformulation of A and B

Chang and Tsai ('99) reformulated A as a family of inner products, one inner product on each $T_\theta\Theta$ as follows.

$$<\gamma_1'(0), \gamma_2'(0)>_A = Cov_\theta[(\left.\frac{d}{dt}\right|_{t=0} \rho(X,\gamma_1(t)))(\left.\frac{d}{ds}\right|_{s=0} \rho(X,\gamma_2(s)))]$$

where γ_1 *and* γ_2 are curves with $\gamma_1(0) = \gamma_2(0) = \theta$.

A is a Riemannian metric on Θ.

Reformulation of A and B

B is reformulated as a family of bilinear form,

one on each $T_\theta\Theta$ by

$$<\gamma_1'(0), \gamma_2'(0)>_B = E_\theta[(\frac{d}{dt}\Big|_{t=0} \rho(X, \gamma_2(t))(\frac{d}{ds}\Big|_{s=0} \log(f(X, \gamma_2(s)))].$$

under some conditions,

$$<\gamma_1'(0), \gamma_2'(0)>_B = -E_\theta[(\frac{\partial^2}{\partial t \partial s}\Big|_{\substack{s=0\\t=0}} \rho(X, \gamma(s,t)))]$$

where $\gamma : R^2 \to \Theta$ satisfies $\gamma(0,0) = \theta$, $\gamma_1(s) = \gamma(s,0)$,

and $\gamma_2(t) = \gamma(0,t)$.

Chang and Rivest Theorem (2001)

PROPOSITION 1. *Suppose the compact Lie group $\mathscr{H}$ is represented on the real vector space $\mathscr{V}$. Write $\mathscr{V} = \oplus \mathscr{V}_i$ as a direct sum of minimally invariant subspaces. Suppose $\langle , \rangle_0$ is an $\mathscr{H}$-invariant positive definite inner product and $\langle , \rangle$ an $\mathscr{H}$-invariant symmetric bilinear form on $\mathscr{V}$. Then:*

(a) *There exist constants c_i such that $\langle , \rangle = c_i \langle , \rangle_0$ on $\mathscr{V}_i$.*

(b) *If $\mathscr{V}_i$ and $\mathscr{V}_j$ are inequivalent as representations of $\mathscr{H}$, they are orthogonal under $\langle , \rangle$ (and $\langle , \rangle_0$).*

Thus, write $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} = \oplus \mathscr{V}_i$ as a direct sum of minimally invariant subspaces and suppose the $\mathscr{V}_i$ are all inequivalent. Then there are constants c_i and d_i such that

$$
\begin{aligned}
\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_A &= \sum_i c_i \langle \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_0 \\
\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_B &= \sum_i d_i < \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_0 \qquad (16) \\
\delta &= \sum_i \boldsymbol{\delta}_i, \qquad \boldsymbol{\delta}_i \in \mathscr{V}_i.
\end{aligned}
$$

This process constructs an asymptotic distribution in $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$, but $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$. Let $\Phi_{\boldsymbol{\theta}_0} : T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} \to \boldsymbol{\Theta}$ be any map such that $\Phi_{\boldsymbol{\theta}_0}(\mathbf{0}) = \boldsymbol{\theta}_0$ and such that the derivative of $\Phi_{\boldsymbol{\theta}_0}$ at $\mathbf{0}$ is the identity map. This latter condition means that if $\mathbf{v} \in T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$, then $\frac{d}{dt}\big|_{t=0} \Phi_{\boldsymbol{\theta}_0}(t\mathbf{v}) = \mathbf{v}$. Brown's theorem becomes the following.

PROPOSITION 2. *Suppose* $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ *is a sample from* $f(\mathbf{x}; \boldsymbol{\theta}_0)$ *and that* $\hat{\boldsymbol{\theta}}$ *minimizes* $\sum_i \rho(\mathbf{X}_i, \boldsymbol{\theta})$. *Let* $\hat{\boldsymbol{\theta}} = \Phi_{\boldsymbol{\theta}_0}(\hat{\mathbf{h}})$ *and* $\hat{\mathbf{h}} = \sum_{i=1}^{i=r} \hat{\mathbf{h}}_i$ *where* $\hat{\mathbf{h}}_i \in \mathscr{V}_i$. *Then the asymptotic distribution of* $n^{1/2}\hat{\mathbf{h}}$ *is multivariate normal with density proportional to*

$$\exp\left(-\frac{n}{2}\sum_i \frac{d_i^2}{c_i}\langle \mathbf{h}_i, \mathbf{h}_i\rangle_0\right), \quad \mathbf{h}_i \in \mathscr{V}_i,\ i = 1, \ldots, r.$$

In particular,

$$n\sum_i \frac{d_i^2}{c_i}\langle \hat{\mathbf{h}}_i, \hat{\mathbf{h}}_i\rangle_0$$

is asymptotically $\chi^2(\dim \boldsymbol{\Theta})$.

calculate c and d

To calculate c and d it suffices to look only at $T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$.

Let $\tilde{h}$ =ϕ_*([0,h]), $\tilde{k}$ =ϕ_*([0,k]) $\in T_{\widetilde{\theta}_0}\boldsymbol{R}P^{p-1}$

where h, k $\in R^{p-1}$. Let γ_1(t) and γ_2(s) be curves in S^{p-1} such that γ_1(0) = γ_2(0)= θ_0, $\gamma_1{}'$=[0,h], and $\gamma_2{}'$=[0,k]. Then if $\widetilde{\gamma_1}$(t) = $\phi(\gamma_1$(t)) and $\widetilde{\gamma_2}$(s) = $\phi(\gamma_2$(s)), we have $\widetilde{\gamma_1}'$ (0)= ϕ_*([0,h])= $\tilde{h}$ and similarly $\widetilde{\gamma_2}'$ (0)= $\tilde{k}$.

By definition

$$
\begin{aligned}
< \tilde{h}, \tilde{k} >_A &= \mathrm{Cov}_{\tilde{\theta}_0}\left[\left(\frac{d}{dt}\Big|_{t=0}\tilde{\rho}(\tilde{x}, \tilde{\gamma}_1(t))\right)\left(\frac{d}{ds}\Big|_{s=0}\tilde{\rho}(\tilde{x}, \tilde{\gamma}_2(s))\right)\right] \\
&= \int_{RP^{p-1}}\left(\frac{d}{dt}\Big|_{t=0}\tilde{\rho}(\tilde{x}, \tilde{\gamma}_1(t))\right)\left(\frac{d}{ds}\Big|_{s=0}\tilde{\rho}(\tilde{x}, \tilde{\gamma}_2(s))\right)\tilde{f}(\tilde{x}, \tilde{\theta}_0) \\
&= 2\int_{S_+^{p-1}}\left(\frac{d}{dt}\Big|_{t=0}\rho_0[(x^t\gamma_1(t))^2]\right)\left(\frac{d}{ds}\Big|_{s=0}\rho_0[(x^t\gamma_2(s))^2]\right)f_0[(x^t\theta_0)^2] \\
&= 4\int_{S^{p-1}}\left(\rho_0'[(x^t\theta_0)^2][0,h]x\right)\left(\rho_0'[(x^t\theta_0)^2]x^t[0,k]^t\right)f(x;\theta_0) \\
&= 4\ [0,h]\ \mathrm{E}_{\theta_0}\left[\left(\rho_0'[(x^t\theta_0)^2]\right)^2 xx^t\right]\ [0,k]^t. \qquad (4)
\end{aligned}
$$

Here we have used (see Lemma 1 of Chang-Rivest)

$$
\mathrm{E}_{\tilde{\theta}_0}\left(\frac{d}{dt}\Big|_{t=0}\tilde{\rho}(\tilde{x}, \tilde{\gamma}_1(t))\right) = 0. \qquad (5)
$$

Let $u = x^t\theta_0$ and write $E_{\theta_0}(xx^t \mid u) = u^2(\theta_0\theta_0^t) + \frac{1-u^2}{p-1}(I - \theta_0\theta_0^t)$. Thus (4) yields

$$< \tilde{h}, \tilde{k} >_A \;=\; \frac{4}{p-1}\int_{S^{p-1}} \left[\left(\rho_0'(u^2)\right)^2 (1-u^2) f_0(u^2)\right] h^t k$$

and hence

$$c \;=\; \frac{4}{p-1}\int_{S^{p-1}} \left[\left(\rho_0'(u^2)\right)^2 (1-u^2) f_0(u^2)\right]. \qquad (6)$$

Similarly, using again (5) and letting $g_0 = \log f_0$,

$$
\begin{aligned}
< \tilde{h}, \tilde{k} >_B &= \mathrm{E}_{\tilde{\theta}_0}\left[\left(\frac{d}{dt}\Big|_{t=0} \log \tilde{f}(\tilde{x}, \tilde{\gamma}_1(t))\right)\left(\frac{d}{ds}\Big|_{s=0} \tilde{\rho}(\tilde{x}, \tilde{\gamma}_2(s))\right)\right] \\
&= 2\int_{S_+^{p-1}} \left(\frac{d}{dt}\Big|_{t=0} g_0[(x^t\gamma_1(t))^2]\right)\left(\frac{d}{ds}\Big|_{s=0} \rho_0[(x^t\gamma_2(s))^2]\right) f_0[(x^t\theta_0)^2] \\
&= 4\,[0,h]\ \mathrm{E}_{\theta_0}\left[g_0^{'}[(x^t\theta_0)^2]\ \rho_0^{'}[(x^t\theta_0)^2]\ xx^t\right]\ [0,k]^t.
\end{aligned}
$$

Hence

$$
d = \frac{4}{p-1}\int_{S^{p-1}} \left[g_0^{'}(u^2)\rho_0^{'}(u^2)(1-u^2)f_0(u^2)\right].
$$

Estimating constants c and d

According to equation (6)

$$c = \frac{4}{p-1}\mathrm{E}_{S^{p-1}}\left[\left(\rho_0'((\mathbf{X}^{\mathbf{t}}\theta)^{\mathbf{2}})\right)^2(1-(\mathbf{X}^{\mathbf{t}}\theta)^{\mathbf{2}})\right].$$

It follows that if $\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_n \in RP^{p-1}$ is a sample from a density $\tilde{f}$ of the form $\tilde{f}(\phi(x); \phi(\theta)) = 2f_0[(x^t\theta)^2]$, then

$$\hat{c} = \frac{4}{n(p-1)}\sum_{i=1}^{n}\left[\left(\rho_0'((\mathbf{X}_{\mathbf{i}}^{\mathbf{t}}\hat{\theta})^{\mathbf{2}})\right)^2(1-(\mathbf{X}_{\mathbf{i}}^{\mathbf{t}}\hat{\theta})^{\mathbf{2}})\right] \qquad (7)$$

is a consistent estimator of c. Here $\mathbf{X_1}, \ldots, \mathbf{X_n} \in \mathbf{S_+^{p-1}}$ satisfy $\phi(\mathbf{X_i}) = \tilde{\mathbf{X}}_{\mathbf{i}}$ and $\phi(\hat{\theta})$, with $\hat{\theta} \in S_+^{p-1}$, is the M-estimate defined by $\tilde{\rho}$.

To develop a sample estimate for d, we use equation (12) of Chang-Rivest.

$$\begin{aligned}
< \tilde{h}, \tilde{k} >_B &= -\mathrm{E}_{\tilde{\theta}_0}\left[\frac{\partial^2}{\partial t \partial s}\bigg|_{(t,s)=(0,0)} \tilde{\rho}(\tilde{x}, \tilde{\gamma}(t,s))\right] \\
&= -2\int_{S_+^{p-1}}\left[\frac{\partial^2}{\partial t \partial s}\bigg|_{(t,s)=(0,0)} \tilde{\rho}(\tilde{x}, \tilde{\gamma}(t,s))\right] f_0[(x^t\theta_0)^2] \\
&= -\mathrm{E}_{\theta_0}\left[\frac{\partial^2}{\partial t \partial s}\bigg|_{(t,s)=(0,0)} \rho(x, \gamma(t,s))\right] \\
&= -\mathrm{E}_{\theta_0}\left[\frac{\partial^2}{\partial t \partial s}\bigg|_{(t,s)=(0,0)} \rho_0[(x^t\gamma(t,s))^2]\right]
\end{aligned}$$

where $\tilde{\gamma}(t,s)$ satisfies $\tilde{\gamma}(t,0) = \tilde{\gamma}_1(t)$ and $\tilde{\gamma}(0,s) = \tilde{\gamma}_2(s)$.

Assume h has unit length, and let $\gamma_1(t) = \cos(t)\theta_0 + \sin(t)h$ and $\gamma(t,s) = \gamma_1(t+s)$. Then

$$
\begin{aligned}
<\tilde{h},\tilde{h}>_B &= -\mathrm{E}_{\theta_0}\left[\frac{\partial^2}{\partial t \partial s}\Big|_{(t,s)=(0,0)} \rho_0[(x^t(\cos(t+s)\theta_0 + \sin(t+s)h))^2]\right] \\
&= -\mathrm{E}_{\theta_0}\left[4h^t\rho_0''(u^2)xx^th - 2\rho_0'(u^2)u\right] \\
&= -\mathrm{E}_{\theta_0}h^t\left[4\rho_0''(u^2)xx^t - 2\rho_0'(u^2)u\right]h,
\end{aligned}
$$

where we have used $h^th = 1$. Thus, using $E_{\theta_0}(xx^t \mid u) = u^2(\theta_0\theta_0^t) + \frac{1-u^2}{p-1}(I - \theta_0\theta_0^t)$

$$
<\tilde{h},\tilde{h}>_B = -\mathrm{E}_{\theta_0}h^t\left[4\rho_0''(u^2)\frac{1-u^2}{p-1} - 2\rho_0'(u^2)u\right]h.
$$

It follows

$$d = -\mathrm{E}_{\theta_0}\left[\frac{4}{p-1}\rho_0''(u^2)(1-u^2) - 2\rho_0'(u^2)u\right]$$

and hence a consistent estimator for d is

$$\hat{d} = -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{4}{p-1}\rho_0''((\mathbf{X}_i^t\hat{\theta})^2)(1-(\mathbf{X}_i^t\hat{\theta})^2) - 2\rho_0'((\mathbf{X}_i^t\hat{\theta})^2)(\mathbf{X}_i^t\hat{\theta})\right]. \qquad (8)$$

Note that the estimates (7) and (8) do not require knowledge of the underlying density f_0.