Data-driven goodness-of-fit tests

Mikhail Langovoy

Institute for Mathematical Stochastics,
Georg-August-University of Göttingen

July 9, 2007

Outline

Outline

Outline

Basic Idea

Constructing good tests is an essential problem of statistics.

Two approaches:

- *naive*: "distance" between theoretical and empirical distribution is proposed as statistic
- *aiming optimality*: construct tests which are asymptotically efficient (Neyman, Le Cam, Wald)

Examples of Distance-Based Statistics

- Kolmogorov - Smirnov:

$$D_n = \sqrt{n}\,\|F_n - F\|_\infty$$

- Cramer - von Mises:

$$\omega_n^2 = n \int_{-\infty}^{\infty} \left(F_n(t) - F_0(t)\right)^2 d\,F_0(t)$$

- many other

About Distance-Based Tests

- These tests works
- Each one is asymptotically optimal only in a very few directions of alternatives

Neyman's Statistic

- hypothesis $H_0 : X \sim U[0,1]$
- $\{\phi_j\}_{j=0}^{\infty}$ orthonormal basis of $L_2([0,1],\lambda)$
- $$N_k = \sum_{j=1}^{k} \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2$$

Neyman's Statistic

- hypothesis $H_0 : X \sim U[0,1]$
- $\{\phi_j\}_{j=0}^{\infty}$ orthonormal basis of $L_2([0,1], \lambda)$
- $$N_k = \sum_{j=1}^{k} \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2$$

Neyman's Statistic

- hypothesis $H_0 : X \sim U[0,1]$
- $\{\phi_j\}_{j=0}^{\infty}$ orthonormal basis of $L_2([0,1], \lambda)$
- $$N_k = \sum_{j=1}^{k} \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2$$

Outline

Idea of Selection Rule.

- model dimension k was known fixed in advance
- important to select the right model dimension
  - incorrect choice can decrease the power of a test
- **Solution**: incorporate the test statistic by some procedure which chooses the right dimension automatically by the data

Idea of Selection Rule.

- model dimension k was known fixed in advance
- important to select the right model dimension
  - incorrect choice can decrease the power of a test
- **Solution**: incorporate the test statistic by some procedure which chooses the right dimension automatically by the data

Idea of Selection Rule.

- model dimension k was known fixed in advance
- important to select the right model dimension
  - incorrect choice can decrease the power of a test
- **Solution**: incorporate the test statistic by some procedure which chooses the right dimension automatically by the data

Bonuses of data-driven score tests

Data-driven score tests are

- asymptotically optimal in an infinite number of directions
- show good overall performance in practice

Outline

Outline

Framework

- $X_1, X_2, \ldots$ random variables, values in a measurable space $\mathbb{X}$
- $X_1, \ldots, X_m$ have joint distribution P_m from the family of distributions $\mathbb{P}_m$ - for every m
- function $\mathcal{F}$ acting from $\otimes_{m=1}^{\infty} \mathbb{P}_m = (\mathbb{P}_1, \mathbb{P}_2, \ldots)$ to a known set Θ
- $\mathcal{F}(P_1, P_2, \ldots) = \theta$

Framework

- $$H_0 : \quad \theta \in \Theta_0 \subset \Theta$$
- $$H_A : \quad \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

- observations $Y_1, \ldots, Y_n$, values in a measurable space $\mathbb{Y}$
- not necessarily on the basis of $X_1, \ldots, X_m$!

Framework

- Θ *any* set
- $Y_1, \ldots, Y_n$ can be dependent or nonidentically distributed
- $\mathbb{Y}$ can be infinite dimensional
- additional assumptions on Y_i's will be imposed when necessary

Outline

Nested families

Definition

- M_k statistical model for specific problem
- Θ_k is the parameter set of M_k
- the family of models M_k for $k = 1, 2, \ldots$ is nested if $\Theta_1 \subseteq \Theta_2 \subseteq \ldots$

- We do not require Θ_k's to be finite dimensional
- We do not require that all Θ_k's are different

Selection rule

T_k *arbitrary* statistic for testing validity of M_k on the basis of $Y_1, \ldots, Y_n$

Penalty

Definition

- nested family of models M_k for $k = 1, \ldots, d(n)$
- $d(n)$ control sequence
- choose $\pi(\cdot, \cdot) : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$
- assume
  - $\pi(1, n) < \pi(2, n) < \ldots < \pi(d(n), n)$ for all n
  - $\pi(j, n) - \pi(1, n) \to \infty$ as $n \to \infty$ for $j = 2, \ldots, d(n)$

Call $\pi(j, n)$ a penalty attributed to jth model M_j and sample size n.

Selection rule

Definition

A selection rule S for the sequence of statistics $\{T_k\}$ is

$$S = \min\{k : 1 \leq k \leq d(n);\ T_k - \pi(k,n) \geq T_j - \pi(j,n),\ j = 1,\ldots,d(n)\}$$

Call T_S a data-driven test statistic for testing validity of the initial model.

Outline

Outline

Main concept

Definition

- observations $Y_1, \ldots, Y_n$, values in a measurable space $\mathbb{Y}$
- k fixed number
- $l = (l_1, \ldots, l_k)$ vector-function
- $l_i : \mathbb{Y} \to \mathbb{R}$ for $i = 1, \ldots, k$ are known Lebesgue measurable

Main concept

Definition

- $$L = \{E_0[I(Y)]^T I(Y)\}^{-1}$$
  - E_0 is with respect to P_0
  - P_0 is the d.f. of some (fixed and known) random variable Y
  - Y has values in $\mathbb{Y}$
- assume
  - $E_0 I(Y) = 0$
  - L is well defined

Main concept

Definition

$$T_k = \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l(Y_j) \right\} L \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l(Y_j) \right\}^T$$

T_k - statistic of Neyman's type (NT-statistic)

Outline

New possibilities

$l_1, \ldots, l_k$ can be

- some score functions
- any other functions, depending on the problem
  - truncated, penalized or partial likelihood
  - possible to use $l_1, \ldots, l_k$ unrelated to any likelihood

Statistical inverse problems

- deconvolution
- building block for many complicated statistical inverse problems

Deconvolution

- instead of X_i one observes Y_i

$$Y_i = X_i + \varepsilon_i$$

- ε_i's are i.i.d. with a known density h
- X_i and ε_i are independent for each i
- $E\,\varepsilon_i = 0$, $0 < E\,\varepsilon^2 < \infty$
- X has a density

Deconvolution

$$H_0: \quad X \text{ has density } f_0$$

- choose for every $k \leq d(n)$ an auxiliary parametric family $\{f_\theta\}$
- $\theta \in \Theta \subseteq \mathbb{R}^k$
- f_0 from this family coincides with f_0 from the null hypothesis H_0
- the true F possibly has no relation to the chosen $\{f_\theta\}$

Deconvolution

$$I(y) = \frac{\frac{\partial}{\partial\theta}\left(\int_{\mathbb{R}} f_\theta(s)\,h(y-s)\,ds\right)\Big|_{\theta=0}}{\int_{\mathbb{R}} f_0(s)\,h(y-s)\,ds}$$

- define U_k as above
- U_k is an NT-statistic

Rank Tests for Independence

- $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. random variables
- distribution function D
- continuous marginal distribution functions F and G

$$H_0 : \quad D(x, y) = F(x)G(y), \quad x, y \in \mathbb{R},$$

Rank Tests for Independence

- b_j denote the $j-$th orthonormal Legendre polynomial $b_1(x) = \sqrt{3}(2x-1)$, $b_2(x) = \sqrt{5}(6x^2 - 6x + 1)$, ...
- R_i the rank of X_i among $X_1, \ldots, X_n$
- S_i the rank of Y_i among $Y_1, \ldots, Y_n$

$$T_k = \sum_{j=1}^{k} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_j\left(\frac{R_i - 1/2}{n}\right) b_j\left(\frac{S_i - 1/2}{n}\right) \right\}^2$$

Outline

Additional assumptions

- $Y_1, Y_2, \ldots$ are identically distributed, with d.f. P
- do not assume $Y_1, Y_2, \ldots$ are independent
- possibly, $X_1, X_2, \ldots$ are dependent and nonidentically distributed

Additional assumptions

- $E_P\, l(Y)$ exists
- $$\frac{1}{n}\sum_{j=1}^{n} l(Y_j) \to E_P\, l(Y) \quad \text{in } P-\text{probability as } n \to \infty$$
- $$n^{-1/2}\sum_{j=1}^{n}(l(Y_j) - E_P\, l(Y)) \to_d \mathcal{N}(0, L^{-1})$$

Consistency condition

$\langle \mathbf{C} \rangle$ there exists integer $K = K(P) \geq 1$ such that

$$E_P\, I_1(Y) = 0, \ldots, E_P\, I_{K-1}(Y) = 0,\ E_P\, I_K = C_P \neq 0$$

Alternative distribution

Theorem

Let the above assumptions holds and

$$\lim_{n\to\infty} \sup_{k\le d(n)} \frac{\pi(k,n)}{n\lambda_k^{(k)}} = 0\,.$$

Then

$$\lim_{n\to\infty} P(S \ge K) = 1.$$

Additional assumptions

suppose there exists a sequence $\{r_n\}_{n=1}^{\infty}$ such that

- $\lim_{n\to\infty} r_n = \infty$
- $$\langle \mathbf{A} \rangle \qquad P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left[I_K(Y_i) - E_P\, I_K(Y_i)\right]\right| \geq y\right) = O\left(\frac{1}{r_n}\right)$$

Alternative distribution

Theorem

Let the above conditions holds and

$$d(n) = o(r_n) \quad as \quad n \to \infty .$$

Then $T_S \to_P \infty$ *as* $n \to \infty$.

Outline

Penalty of proper weight

- $\{T_k\}$ sequence of NT-statistics
- S selection rule
- $\pi(k, n)$ penalty

Penalty of proper weight

Definition

$\pi(k,n)$ is of proper weight, if there exists real sequences $\{s(k,n)\}_{k,n=1}^{\infty}$, $\{t(k,n)\}_{k,n=1}^{\infty}$, such that

- $$\lim_{n\to\infty} \sup_{k\le u_n} \frac{s(k,n)}{n\lambda_k^{(k)}} = 0\,,$$

  $\{u_n\}_{n=1}^{\infty}$ real sequence, $\lim_{n\to\infty} u_n = \infty$

- $\lim_{n\to\infty} t(k,n) = \infty$ for every $k \ge 2$

  $\lim_{k\to\infty} t(k,n) = \infty$ for every fixed n

Penalty of proper weight

Definition

- $s(k,n) \le \pi(k,n) - \pi(1,n) \le t(k,n)$ for all k, n
- $$\lim_{n\to\infty} \sup_{k \le m_n} \frac{\pi(k,n)}{n\lambda_k^{(k)}} = 0\,,$$

  $\{m_n\}_{n=1}^{\infty}$ real sequence, $\lim_{n\to\infty} m_n = \infty$.

Notation

define for $l = (l_1, \ldots, l_k)$

- $$\bar{l}_j := \frac{1}{n} \sum_{i=1}^{n} l_j(Y_i)$$
- $$\bar{l} := (\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_k)$$
- $$Q_k(\bar{l}) = (\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_k)\, L\, (\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_k)^T$$

Proper majorant

Definition

- statistic T_k, selection rule S, penalty of proper weight
- Assume there exists a Lebesgue measurable $\varphi(\cdot, \cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$
  - monotonically decreasing in the second argument
  - monotonically nondecreasing in the first argument
  - and

Proper majorant

Definition

- (B2) for every $\varepsilon > 0$ there exists $K = K_\varepsilon$ such that for $n > n(\varepsilon)$

$$\sum_{k=K_\varepsilon}^{u_n} \varphi(k; s(k,n)) < \varepsilon$$

- (B) for all $k \geq 1$ and $y \in [s(k,n); t(k,n)]$

$$P_0\left(n\, Q_k(\bar{l}) \geq y\right) \leq \varphi(k; y)$$

We call φ a proper majorant for (large deviations of) T_k.

Null distribution

Theorem

- $\{T_k\}$ *sequence of NT-statistics*
- *S selection rule, with penalty of proper weight*
- *large deviations of* T_k *are properly majorated*
- $$d(n) \leq \min\{u_n, m_n\}.$$

Then $S = O_{P_0}(1)$ *and* $T_S = O_{P_0}(1)$.

Consistency theorem

Theorem

- $\{T_k\}$ *sequence of NT-statistics*
- *S selection rule, with penalty of proper weight*
- *the above regularity assumptions are satisfied*
- $d(n) = o(r_n)$, $d(n) \leq \min\{u_n, m_n\}$.

Then T_S is consistent against any (fixed) alternative P satisfying (C).

Outline

Outline

Definition

- The notion of NT-statistics is helpful if the null hypothesis is simple
- for composite hypotheses, the concept needs to be modified

Definition

Definition

- observations $Y_1, \ldots, Y_n$, values in a measurable space $\mathbb{Y}$
- for simplicity, identically distributed
- k fixed number
- $I = (I_1, \ldots, I_k)$ vector-function
- $I_i : \mathbb{Y} \to \mathbb{R}$ for $i = 1, \ldots, k$ (maybe unknown) Lebesgue measurable
- $$L^{(0)} = \{E_0[I(Y)]^T I(Y)\}^{-1}$$
- P_0 (possibly unknown) d.f. of Y under the null hypothesis

Definition

- assume $E_0 l(Y) = 0$
- $L^{(0)}$ is well-defined
- L_k denote for every k a $k \times k$ symmetric positive definite matrix with finite elements
- $$\left\| L_k - L^{(0)} \right\| = o_{P_0}(1)$$

Sufficiently good estimators

Definition

$I_1^*, \ldots, I_n^*$ sufficiently good estimators of $I(Y_1), \ldots, I(Y_n)$ w.r.t. P_0, if

for every $\varepsilon > 0$

$$P_0^n\left(\frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n}(I_j^* - I(Y_j))\right\| \geq \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty$$

Definition of GNT-statistics

Definition

$$GT_k = \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l_j^* \right\} L_k \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l_j^* \right\}^T$$

GT_k generalized statistic of Neyman's type (GNT-statistic)

GT_S data-driven GNT-statistic

Outline

Null distribution of GNT-statistics

Theorem

- $\{GT_k\}$ *sequence of GNT-statistics*
- *S selection rule, with penalty of proper weight (for* R_k*)*
- *large deviations of* GT_k *are properly majorated*

  $d(n) \leq \min\{u_n, m_n\}$.

Then under the null hypothesis $S = O_{P_0}(1)$ *and* $GT_S = O_{P_0}(1)$.

Possible additional assumptions

- $$\langle \mathbf{C1} \rangle \qquad \left\| L - L^{(0)} \right\| = o_P(1)$$

- $l_1^*, \ldots, l_n^*$ are sufficiently good estimators of $l(Y_1), \ldots, l(Y_n)$ with respect to P

- i.e., for every $\varepsilon > 0$

$$P^n \left(\frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{n} (l_j^* - l(Y_j)) \right\| \geq \varepsilon \right) \to 0 \quad \text{as} \quad n \to \infty .$$

Possible additional assumptions

- $$\langle \mathbf{C1} \rangle \qquad \left\| L - L^{(0)} \right\| = o_P(1)$$

- $l_1^*, \ldots, l_n^*$ are sufficiently good estimators of $I(Y_1), \ldots, I(Y_n)$ with respect to P

- i.e., for every $\varepsilon > 0$

$$P^n \left(\frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{n} (l_j^* - I(Y_j)) \right\| \geq \varepsilon \right) \to 0 \quad \text{as} \quad n \to \infty .$$

Consistency theorem for GNT-statistics

Theorem

- $\{GT_k\}$ *sequence of GNT-statistics*
- S *selection rule*
- $d(n) = o(r_n)$.

Under the above assumptions, T_S *is consistent against any (fixed) alternative distribution* P *satisfying* $\langle C \rangle$ *and* $\langle C1 \rangle$.

Outline

- SNT-statistics
- Selection rule: examples
- Application of NT-statistics
- GNT-statistics: examples

Definition of SNT-statistics

Definition

Consider statistic of the form

$$T_k = \sum_{j=1}^{k} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_j(Y_i) \right\}^2$$

- n is the number of available observations $Y_1, \ldots, Y_n$
- $l_i : \mathbb{Y} \to \mathbb{R}$ are known Lebesgue measurable

We call T_k the *simplified statistic of Neyman's type* (SNT-statistic).

Partial likelihood (Cox)

- generalizes conditional and marginal likelihood
- applications include inference in stochastic processes

Partial likelihood

- random variable Y, density $f_Y(y;\theta)$
- let Y be transformed into the sequence

$$(X_1, S_1, X_2, S_2, \ldots, X_m, S_m)$$

- components may themselves be vectors

Partial likelihood

- full likelihood

$$\prod_{j=1}^{m} f_{X_j|X^{(j-1)},S^{(j-1)}}(x_j|x^{(j-1)},s^{(j-1)};\theta) \prod_{j=1}^{m} f_{S_j|X^{(j)},S^{(j-1)}}(s_j|x^{(j)},s^{(j-1)};\theta)$$

$x^{(j)} = (x_1,\ldots,x_j)$ and $s^{(j)} = (s_1,\ldots,s_j)$

- second product is *partial likelihood*

Partial likelihood

- for simplicity θ is real
- $H_0 : \theta = \theta_0$
- define for $j = 1, \ldots, m$

$$t_j = \left. \frac{\partial \log f_{S_j|X^{(j)},S^{(j-1)}}(s_j|x^{(j)}, s^{(j-1)}; \theta)}{\partial \theta} \right|_{\theta=\theta_0}$$

and $\sigma_j^2 := var(t_j)$

Partial likelihood

- define $l_j := t_j / \sigma_j$
- SNT-statistic

$$PL_m = \sum_{j=1}^{m} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_j \right\}^2$$

Outline

- SNT-statistics
- Selection rule: examples
- Application of NT-statistics
- GNT-statistics: examples

Partial likelihood

- PL_m depends on the number m of components
- suppose Y can be transformed into sequences (X_1, S_1), or (X_1, S_1, X_2, S_2), or even $(X_1, S_1, X_2, S_2, \ldots, X_m, S_m)$ for any m
- which m is the best choice?
- PL_S chooses the reasonable number of components automatically

Gaussian model selection (Birgé and Massart)

- framework of Gaussian linear processes
- includes
  - Gaussian regression with fixed design
  - Gaussian sequences
  - model of Ibragimov and Has'minskii

Gaussian model selection

$\mathbb{S}$ linear subspace of some Hilbert space $\mathbb{H}$

Definition

Gaussian linear process on $\mathbb{S}$ with mean $s \in \mathbb{H}$ and variance ε^2 any process Y indexed by $\mathbb{S}$ of the form

$$Y(t) = \langle s, t \rangle + \varepsilon Z(t),$$

for all $t \in \mathbb{S}$

Z linear isonormal process indexed by $\mathbb{S}$ (i.e. Z is a centered and linear Gaussian process with covariance structure $E[Z(t)Z(u)] = \langle t, u \rangle$)

Birgé and Massart considered estimation of s in this model.

Gaussian model selection

- S finite dimensional subspace of $\mathbb{S}$
- $\gamma(t) = \|t\|^2 - 2Y(t)$
- define the projection estimator on S to be the minimizer of $\gamma(t)$ with respect to $t \in S$

Gaussian model selection

Given

- a finite or countable family $\{S_m\}_{m\in\mathcal{M}}$ of finite dimensional linear subspaces of S
- the corresponding family of projection estimators $\widehat{s}_m$, built for the same realization of process Y
- a nonnegative function *pen* defined on $\mathcal{M}$

estimate s by $\widetilde{s} = \widehat{s}_{\widehat{m}}$, where $\widehat{m}$ is any minimizer with respect to $m \in \mathcal{M}$ of the penalized criterion

$$crit(m) = -\|\widehat{s}_m\|^2 + pen(m) = \gamma(\widehat{s}_m) + pen(m)$$

Gaussian model selection

- $\gamma(t)$ is the least squares criterion
- $\widehat{s}_m$ is the least squares estimator of s, which is in this case the maximum likelihood estimator
- $\|\widehat{s}_m\|^2$ is the Neyman score for testing $s = 0$ within this model

Gaussian model selection

- *pen* is a selection rule in our terminology
- $\|\widehat{s}_{\widehat{m}}\|^2$ is the data-driven SNT-statistics
- by the Consistency Theorem, $\|\widehat{s}_{\widehat{m}}\|^2$ can be used for testing $s = 0$ and has a good range of consistency

Partial likelihood

- possible to define PL_m for θ multidimensional or even infinite-dimensional
- under additional regularity assumptions $E(t_j) = 0$
- PL_m NT-statistic, but not SNT-statistic

Example

Lemma

Let $l_K(Y_i)'s$ be bounded i.i.d. random variables with finite variance σ^2. Then $\langle A \rangle$ is satisfied with $r_n = \exp(ny^2/2\sigma)$.

Outline

- SNT-statistics
- Selection rule: examples
- Application of NT-statistics
- GNT-statistics: examples

Application to SNT-statistics

Theorem

- $Y_1, \ldots, Y_n$ *i.i.d.*
- $\{T_k\}$ *family of SNT-statistics, S selection rule*
- $E\, l(Y_1) = 0$
- $(l_1(Y_i), \ldots, l_k(Y_i))$ *has unit covariance matrix for every k*
- $\|(l_1(Y_1), \ldots, l_k(Y_1))\|_k \leq M(k)$ *a.e.*

Application to SNT-statistics

Theorem

- $\pi(k,n) - \pi(1,n) \geq 2k$ *for all* $k \geq 2$
- $$\lim_{n\to\infty} \frac{M(d(n))\,\pi(d(n),n)}{\sqrt{n}} = 0.$$

Then $S = O_{P_0}(1)$ *and* $T_S = O_{P_0}(1)$.

Neyman's statistics

Theorem

- *T_S Neyman's smooth data-driven test statistic*
- *$\pi(k,n) - \pi(1,n) \geq 2k$ for all $k \geq 2$ and for all $k \leq d(n)$*
- $$\lim_{n\to\infty} \frac{d(n)\pi(d(n),n)}{\sqrt{n}} = 0.$$

Then $S = O_{P_0}(1)$ and $T_S = O_{P_0}(1)$.

Rank tests for independence

Theorem

- $d(n) = o\left(\left\{\frac{n}{\log n}\right\}^{1/10}\right)$
- $\mathbb{P}$ *alternative*
- F *and* G *the marginal distribution functions of* X *and* Y *under* $\mathbb{P}$
- *for some* j

$$E_{\mathbb{P}}\, b_j(F(X))b_j(G(Y)) \neq 0.$$

If $d(n) \to \infty$, *then* $T_S \to \infty$ *as* $n \to \infty$.

Rank tests for independence

- condition $\langle C \rangle$: for some j

$$E_{\mathbb{P}}\, b_j\left(\frac{R_1 - 1/2}{n}\right) b_j\left(\frac{S_1 - 1/2}{n}\right) \neq 0$$

- for continuous F and G this is asymptotically equivalent to

$$E_{\mathbb{P}}\, b_j(F(X)) b_j(G(Y)) \neq 0.$$

Outline

- SNT-statistics
- Selection rule: examples
- Application of NT-statistics
- GNT-statistics: examples

Smooth test for composite hypotheses

- $X_1, \ldots X_n$ i.i.d. with density $f(x)$
- $H_0: \quad f(x) \in \{f(x;\beta), \beta \in \mathcal{B}\}$
- $\mathcal{B} \subset \mathbb{R}^q$
- $\{f(x;\beta), \beta \in \mathcal{B}\}$ given family of densities

Smooth test for composite hypotheses

- F distribution function corresponding to f
- $$Y_n(\beta) = n^{-1} \sum_{i=1}^{n} (\phi_1(F(X_i;\beta)), \ldots, \phi_j(F(X_i;\beta)))^T$$
- j depends on the context
- I the $k \times k$ identity matrix

Smooth test for composite hypotheses

- $$I_\beta = \left\{ -E_\beta \frac{\partial}{\partial \beta_t} \phi_j(F(X_i;\beta)) \right\}_{t=1,\ldots,q;\, j=1,\ldots,k}$$
- $$I_{\beta\beta} = \left\{ -E_\beta \frac{\partial^2}{\partial \beta_t \partial \beta_u} \log f(X;\beta) \right\}_{t=1,\ldots,q;\, u=1,\ldots,q}$$
- $$R(\beta) = I_\beta^T (I_{\beta\beta} - I_\beta I_\beta^T) I_\beta$$

Smooth test for composite hypotheses

- $\widehat{\beta}$ maximum likelihood estimator of β under H_0
- the score statistic

$$W_k(\widehat{\beta}) = n\, Y_n^T(\widehat{\beta})\{I + R(\widehat{\beta})\}\, Y_n(\widehat{\beta})$$

- in a regular situation (Cox and Hinkley), $W_k(\widehat{\beta})$ is a GNT-statistic