Washington State University

Xiongzhi Chen

Email: xiongzhi.chen@wsu.edu

Department of Mathematics and Statistics

Proportion estimators, Lebesgue-Stieltjes integral equations, and concentration of measures

Outline

- Background on proportion estimation
- Brief review on existing methods
- The new method
- Summary

Large-scale multiple hypotheses testing

- Identify among hundreds of drugs those that may have a specific side effect
- Identify among thousands of genes those that might be associated with a disease
- Identify among millions of SNPs those that might be associated with the yield of a crop

Four basic questions

Signals are entities that provide scientific information or that a scientist is searching for. Given a data set, one asks 4 basic questions:

1. Are there any signals? (Testing)
2. If there are signals, how many signals? (Testing)
3. If there are signals, where are they? (Testing)
4. If there are signals, what are their magnitudes? (Estimation)

The proportion of false nulls

- Convention: "a signal" is identified with "a false null", and "a noise" identified with "a true null"

- Among m null hypotheses, the proportion of false nulls, i.e., the signal proportion, is defined as

$$\pi_{1,m} = \frac{\text{Number of false nulls}}{m}$$

- Example: in gene-disease association study, $\pi_{1,m}$ is the ratio of the number of disease-contributing genes to the total number of genes under investigation

Paradigms of sparsity

- Dense regime: $\liminf_{m\to\infty} \pi_{1,m} > 0$, i.e., there is always a positive proportion of signals
- Moderately sparse regime: $\pi_{1,m} = O\left(m^{-s}\right)$ for $0 < s < 0.5$
- Critically sparse regime: $\pi_{1,m} = O\left(m^{-0.5}\right)$
- Very sparse regime: $\pi_{1,m} = O\left(m^{-s}\right)$ for $0.5 < s < 1$

Formulation of estimation problem

- Data: $\{z_i\}_{i=1}^m$ rv's each with mean or median μ_i
- Hypotheses: $H_{i0} : \mu_i = \mu_0$ vs $H_{i1} : \mu_i \neq \mu_0$
- Indices of signals: $I_{1,m} = \{1 \leq i \leq m : \mu_i \neq \mu_0\}$
- Signal proportion: $\pi_{1,m} = m_1 m^{-1}$, where $m_1 = |I_{1,m}|$
- Target: to consistently estimate $\pi_{1,m}$

A brief review on existing approaches

Four approaches to proportion estimation:

1. Hard thresholding: H_{i0} is false when $|z_i| \geq t$ for some $t > 0$
2. Deconvolving two-component mixture model: $\{z_i\}_{i=1}^{m}$ identically distributed as

$$f(x) = \left(1 - \pi_{1,m}\right) f_0(x) + \pi_{1,m} f_1(x)$$

3. Bounding an excess of uniform empirical process
4. Fourier transform, implemented for Gaussian family or mixtures with a Gaussian component

Disadvantages of the four approaches:

- Only Approaches 3 and 4 are consistent when $\pi_{1,m}$ tends to zero
- Approach 4 is not applicable when CDFs of rv's do not consist at least one component from a location-shift family
- Consistency of Approach 3 requires CDFs to be absolutely continuous
- Consistency of Approaches 1 and 2 require restrictive modelling assumptions (e.g., z_i's corresponding to false nulls are identically distributed)

The strategy: set-up

- Let $\mathbf{z} = (z_1, \ldots, z_m)$ and $\boldsymbol{\mu} = (\mu_1, ..., \mu_m)$
- Denote by F_{μ_i} the CDF of z_i such that $F_{\mu_i} \in \mathscr{F} = \{F_\mu : \mu \in U\}$, where U has a non-empty interior
- Each F_μ is uniquely determined by μ
- Recall: $H_{i0} : \mu_i = \mu_0 \quad \text{vs} \quad H_{i1} : \mu_i \neq \mu_0$

The strategy: constructing "Phase function"

- For each $\mu \in U$, approximate $1_{\{\mu \neq \mu_0\}}$ by $\psi(t, \mu; \mu_0)$, $t \in \mathbb{R}$ such that

$$\lim_{t\to\infty} \psi(t, \mu_0; \mu_0) = 1 \text{ and } \lim_{t\to\infty} \psi(t, \mu; \mu_0) = 0 \text{ for } \mu \neq \mu_0 \quad (1)$$

- Define "phase function"

$$\varphi_m(t, \boldsymbol{\mu}) = \frac{1}{m} \sum_{i=1}^{m} [1 - \psi(t, \mu_i; \mu_0)] \quad (2)$$

- The phase function: $\lim_{t\to\infty} \varphi_m(t, \boldsymbol{\mu}) = \pi_{1,m}$ for any fixed m and $\boldsymbol{\mu}$, i.e., the "Oracle"

$$\Lambda_m(\boldsymbol{\mu}) = \lim_{t\to\infty} \varphi_m(t, \boldsymbol{\mu}) \quad (3)$$

- Construct $K:\mathbb{R}^2 \to \mathbb{R}$ that is independent of $\mu \neq \mu_0$ and solves the Lebesgue-Stieltjes integral equation

$$\psi\left(t,\mu;\mu_0\right) = \int K\left(t,x;\mu_0\right) dF_\mu\left(x\right) \tag{4}$$

- Define "empirical phase function"

$$\hat{\varphi}_m\left(t,\mathbf{z}\right) = \frac{1}{m}\sum_{i=1}^{m}\left[1-K\left(t,z_i;\mu_0\right)\right] \tag{5}$$

- The empirical phase function: for any fixed m, t and $\boldsymbol{\mu}$

$$\mathbb{E}\left[\hat{\varphi}_m\left(t,\mathbf{z}\right)\right] = \varphi_m\left(t,\boldsymbol{\mu}\right) \tag{6}$$

The strategy: intuition for consistency

- Recall "phase function" $\varphi_m(t,\boldsymbol{\mu})$ and $\lim_{t\to\infty}\varphi_m(t,\boldsymbol{\mu}) = \pi_{1,m}$
- Recall "empirical phase function" $\hat{\varphi}_m(t,\mathbf{z})$ and

$$\mathbb{E}\left[\hat{\varphi}_m(t,\mathbf{z})\right] = \varphi_m(t,\boldsymbol{\mu})$$

- Intuition: $\hat{\varphi}_m(t,\mathbf{z})$ will consistently estimate $\pi_{1,m}$, i.e.,

$$\Pr\left(\left|\frac{\hat{\varphi}_m(t_m,\mathbf{z})}{\pi_{1,m}} - 1\right| \to 0\right) \to 1 \quad \text{as } m\to\infty \tag{7}$$

if

$$e_m(t_m) = \left|\hat{\varphi}_m(t_m,\mathbf{z}) - \varphi_m(t_m,\boldsymbol{\mu})\right| \to 0 \text{ at appropriate speed}$$

The strategy: main difficulties in implementation

- Phase functions may not exist, i.e., no solution in ψ to

$$\lim_{t\to\infty} \psi\left(t, \mu_0; \mu_0\right) = 1 \text{ and } \lim_{t\to\infty} \psi\left(t, \mu; \mu_0\right) = 0 \text{ for } \mu \neq \mu_0$$

  or in K (independent of $\mu \neq \mu_0$) to

$$\psi\left(t, \mu; \mu_0\right) = \int K\left(t, x; \mu_0\right) dF_\mu(x)$$

- The oscillation

$$e_m(t_m) = \left|\hat{\varphi}_m(t_m, \mathbf{z}) - \varphi_m\left(t_m, \boldsymbol{\mu}\right)\right|$$

  may not converge to 0 when K is not a Lipschitz or bounded transform of $\{z_i\}_{i=1}^{m}$

The strategy: summary of results

When $\{z_i\}_{i=1}^m$ are mutually independent, uniformly consistent estimators $\hat{\varphi}_m(t, \mathbf{z})$ are constructed respectively for the estimation problem in the settings where

- Random variables have Riemann-Lebesgue type characteristic functions (RL Type CFs)
- Random variables have CDFs that are members of natural exponential families with support $\mathbb{N}$
- Random variables have CDFs that are members of natural exponential families with separable moments

- Recall $\mathscr{F} = \{F_\mu : \mu \in U\}$ and let $\hat{F}_\mu(t) = \int e^{\iota tx} dF_\mu(x)$ be the CF of F_μ where $\iota = \sqrt{-1}$
- Let r_μ be the modulus of $\hat{F}_\mu$, so that $\hat{F}_\mu = r_\mu e^{\iota h_\mu}$

Definition (Riemann-Lebesgue type CF)

If $\{t \in \mathbb{R} : \hat{F}_{\mu_0}(t) = 0\} = \varnothing$, $\sup_{t\in\mathbb{R}} \frac{r_\mu(t)}{r_{\mu_0}(t)} < \infty$ for each $\mu \in U \setminus \{\mu_0\}$, and

$$\lim_{t\to\infty} \frac{1}{t} \int_{[-t,t]} \frac{\hat{F}_\mu(y)}{\hat{F}_{\mu_0}(y)} dy = 0, \tag{8}$$

then $\hat{\mathscr{F}} = \{\hat{F}_\mu : \mu \in U\}$ is said to be of “Riemann-Lebesgue type (RL type)” at μ_0 on U.

- Let $\omega : [-1,1] \to \mathbb{R}_+$ be bounded and Lebesgue-integrates to 1

Theorem (RVs with RL Type CFs)

Let $\hat{\mathscr{F}}$ be of RL type and define $K : \mathbb{R}^2 \to \mathbb{R}$ as

$$K\left(t, x; \mu_0\right) = \int_{[-1,1]} \frac{\omega\left(s\right) \cos\left(tsx - h_{\mu_0}\left(ts\right)\right)}{r_{\mu_0}\left(ts\right)} ds. \tag{9}$$

Then

$$\psi\left(t, \mu; \mu_0\right) = \int_{[-1,1]} \omega\left(s\right) \frac{r_\mu\left(ts\right)}{r_{\mu_0}\left(ts\right)} \cos\left(h_\mu\left(ts\right) - h_{\mu_0}\left(ts\right)\right) ds. \tag{10}$$

Construction I: application to location-shift families

- $\mathscr{F} = \{F_\mu : \mu \in U\}$ is a location-shift family if and only if $z + \mu'$ has CDF $F_{\mu+\mu'}$ whenever z has CDF F_μ for $\mu, \mu + \mu' \in U$

Corollary (Construction I for Location-shift families)

If $\mathscr{F}$ is a location-shift family for which $\{t \in \mathbb{R} : \hat{F}_{\mu_0}(t) = 0\} = \varnothing$, then

$$\psi(t, \mu; \mu_0) = \int K(t, y + (\mu - \mu_0); \mu_0)\, dF_{\mu_0}(y)$$
$$= \int_{[-1,1]} \omega(s) \cos(ts(\mu - \mu_0))\, ds.$$

- Gaussian family $\mathsf{Normal}\left(\mu,\sigma^2\right)$ with mean μ and std dev $\sigma > 0$, for which

$$K\left(t,x;\mu_0\right) = \int_{[-1,1]} \exp\left(2^{-1}t^2s^2\sigma^2\right)\omega\left(s\right)\exp\left(\iota ts\left(x-\mu_0\right)\right)ds$$

- Laplace family $\mathsf{Laplace}\left(\mu,2\sigma^2\right)$ with mean μ and std dev $\sqrt{2}\sigma > 0$, for which

$$K\left(t,x;\mu_0\right) = \int_{[-1,1]} \left(1+\sigma^2t^2s^2\right)\omega\left(s\right)\exp\left(\iota ts\left(x-\mu_0\right)\right)ds$$

- Cauchy family $\mathsf{Cauchy}\left(\mu,\sigma\right)$ with median μ and scale parameter $\sigma > 0$, for which

$$K\left(t,x;\mu_0\right) = \int_{[-1,1]} \exp\left(\sigma\left|ts\right|\right)\omega\left(s\right)\exp\left(\iota ts\left(x-\mu_0\right)\right)ds$$

Definition (Uniform consistency class)

Given a family $\mathscr{F}$, the sequence of sets $\mathscr{Q}_m\left(\boldsymbol{\mu}, t; \mathscr{F}\right) \subseteq \mathbb{R}^m \times \mathbb{R}$ for each $m \in \mathbb{N}_+$ is called a "uniform consistency class" for the estimator $\hat{\varphi}_m(t, \mathbf{z})$ if

$$\Pr\left(\sup_{\boldsymbol{\mu} \in \mathscr{Q}_m(\boldsymbol{\mu}, t; \mathscr{F})} \left|\pi_{1,m}^{-1} \sup_{t \in \mathscr{Q}_m(\boldsymbol{\mu}, t; \mathscr{F})} \hat{\varphi}_m(t, \mathbf{z}) - 1\right| \to 0\right) \to 1.$$

- Define

$$\mathscr{B}_m\left(\rho\right) = \left\{\boldsymbol{\mu} \in \mathbb{R}^m : m^{-1} \sum_{i=1}^m \left|\mu_i - \mu_0\right| \le \rho\right\} \text{ for some } \rho > 0$$

and $u_m = \min\left\{\left|\mu_j - \mu_0\right| : \mu_j \neq \mu_0\right\}$

Theorem (Uniform consistency for Construction I)

If $\mathscr{F}$ is a "good" location-shift family, then a uniform consistency class is

$$\mathscr{Q}_m(\boldsymbol{\mu}, t; \mathscr{F}) = \left\{ \begin{array}{l} R_m(\rho) = O\left(m^{\vartheta'}\right), \tau_m \le \gamma_m, u_m \ge \frac{\log\log m}{\gamma'' \tau_m}, \\ t \in [0, \tau_m], \lim_{m \to \infty} \pi_{1,m}^{-1} \Upsilon(q, \tau_m, \gamma_m, r_{\mu_0}) = 0 \end{array} \right\},$$

where $R_m(\rho) \sim \rho$, $\gamma_m = \gamma' \log m$

$$\Upsilon(q, \tau_m, \gamma_m, r_{\mu_0}) = \frac{2\|\omega\|_\infty \sqrt{2q\gamma_m}}{\sqrt{m}} \sup_{t \in [0, \tau_m]} \int_{[0,1]} \frac{ds}{r_{\mu_0}(ts)}.$$

Moreover, for all sufficiently large m, with probability at least $1 - o(1)$*,*

$$\sup_{\boldsymbol{\mu} \in \mathscr{B}_m(\rho)} \sup_{t \in [0, \tau_m]} \left| \hat{\varphi}_m(t, \mathbf{z}) - \varphi_m(t, \boldsymbol{\mu}) \right| \le \Upsilon(q, \tau_m, \gamma_m, r_{\mu_0}).$$

Construction I: interpretation of uniform consistency class

For location-shift families, recall r_{μ_0} as the modulus of CF $\hat{F}_{\mu_0}$ for CDF F_{μ_0} and

$$\Upsilon\left(q,\tau_m,\gamma_m,r_{\mu_0}\right)=\frac{2\left\|\omega\right\|_\infty\sqrt{2q\gamma_m}}{\sqrt{m}}\sup_{t\in[0,\tau_m]}\int_{[0,1]}\frac{ds}{r_{\mu_0}(ts)}$$

- The speed of convergence and uniform consistency class for the estimator $\hat{\varphi}_m(t,\mathbf{z})$ is mainly determined by r_{μ_0} (not by r_μ with $\mu\neq\mu_0$) and

$$u_m=\min\left\{\left|\mu_j-\mu_0\right|:\mu_j\neq\mu_0\right\}$$

- Uniform consistency can be achieved when $\{z_i\}_{i=1}^m$ do not have bounded first-order absolute moment (e.g., Cauchy family)
- It is unlikely that $\hat{\varphi}_m(t,\mathbf{z})$ can be consistent for $\pi_{1,m}=O\left(m^{-0.5}\right)$

- Let β be a positive non-Dirac Radon measure on $\mathbb{R}$
- Let $L(\theta) = \int e^{x\theta} \beta(dx)$ for $\theta \in \mathbb{R}$ and Θ be the maximal open set containing θ such that $L(\theta) < \infty$
- Assume Θ is not empty and let $\kappa(\theta) = \log L(\theta)$. Then

$$\mathscr{F} = \left\{ G_\theta : G_\theta(dx) = \exp(\theta x - \kappa(\theta)) \beta(dx), \theta \in \Theta \right\}$$

  forms an NEF with respect to the basis β
- Fact: Θ is convex if it has a non-empty interior; L is analytic on the strip $A_\Theta = \{z \in \mathbb{C} : \Re(z) \in \Theta\}$
- Fact: $\mathscr{F}$ can be equivalently characterized as $\mathscr{F} = \left\{ F_\mu : \mu \in U \right\}$, where $\mu : \Theta \to U$ is the mean function with $U = \mu(\Theta)$

Constructions II and III: notational set-up

- μ_0 is identified with θ_0, and $\mu = \mu(\theta)$ for $\theta \in \Theta$
- Reuse ψ but take it as a function of t and θ
- Reuse K but take it as a function of t and θ, such that

$$\psi(t,\theta;\theta_0) = \int K(t,x;\theta_0)\, dG_\theta(x) \text{ for } G_\theta \in \mathscr{F}.$$

- Let $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)$, the phase function

$$\varphi_m(t,\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\left[1-\psi(t,\theta_i;\theta_0)\right]$$

  and the empirical phase function

$$\hat{\varphi}_m(t,\mathbf{z}) = \frac{1}{m}\sum_{i=1}^{m}[1-K(t,z_i;\theta_0)]$$

Outside families with Riemann-Lebesgue type CFs

- Fourier transform used to construct K is no longer applicable
- K does not always exist as a solution to

$$\psi(t,\theta;\theta_0) = \int K(t,x;\theta_0)\, dG_\theta(x) \text{ for } G_\theta \in \mathscr{F}$$

- Even if K does exist, the transform induced by K on $\{z_i\}_{i=1}^{m}$ is no longer Lipschitz or bounded, which causes great difficulty in deriving concentration inequalities for the oscillation

$$e_m(t_m) = \left|\hat{\varphi}_m(t_m, \mathbf{z}) - \varphi_m\left(t_m, \boldsymbol{\mu}\right)\right|$$

- Special functions and their asymptotics are needed since NEFs are often related to these functions

Construction II: discrete NEFs with infinite supports

Assume the basis β for $\mathscr{F}$ is discrete with support $\mathbb{N}$, i.e., there exists a positive sequence $\{c_k\}_{k\geq 0}$ such that

$$\beta = \sum_{k=0}^{\infty} c_k \delta_k$$

- The power series $H(z) = \sum_{k=0}^{\infty} c_k z^k$ with $z \in \mathbb{C}$ must have a positive radius of convergence R_H
- h is the generating function (GF) of β
- If β is a probability measure, then $(-\infty, 0] \subseteq \Theta$ and $R_H \geq 1$, and vice versa

Theorem (Construction for discrete NEF with support $\mathbb{N}$)

Let $\mathscr{F}$ be an NEF generated by β with support $\mathbb{N}$. For $x \in \mathbb{N}$ and $t \in \mathbb{R}$ set

$$K(t,x;\theta_0) = H\left(e^{\theta_0}\right)\int_{[-1,1]} \frac{(ts)^x \cos\left(\frac{\pi x}{2} - tse^{\theta_0}\right)}{H^{(x)}(0)}\omega(s)\,ds. \tag{11}$$

Then

$$\begin{aligned}\psi(t,\theta;\theta_0) &= \int K(t,x;\theta_0)\,dG_\theta(x)\\ &= \frac{H\left(e^{\theta_0}\right)}{H\left(e^{\theta}\right)}\int_{[-1,1]} \cos\left(st\left(e^{\theta} - e^{\theta_0}\right)\right)\omega(s)\,ds. \end{aligned} \tag{12}$$

- Poisson family $\mathsf{Poissson}\left(\mu\right)$ with mean $\mu > 0$. The basis

$$\beta = \sum_{k=0}^{\infty} \frac{1}{k!}\delta_k \text{ with } L(\theta) = \exp\left(e^{\theta}\right) \text{ for } \theta \in \mathbb{R},$$

$\mu(\theta) = e^{\theta}$, $H(z) = e^{z}$ with $R_H = \infty$ and $H^{(k)}(0) = 1$ for all $k \in \mathbb{N}$

- Negative Binomial family $\mathsf{NegBinomial}(\theta, n)$ with $\theta < 0$ and $n \in \mathbb{N}_+$. The basis is

$$\beta = \sum_{k=0}^{\infty} \frac{c_k^*}{k!}\delta_k \text{ with } c_k^* = \frac{(k+n-1)!}{(n-1)!} \text{ for } k \in \mathbb{N},$$

$L(\theta) = \left(1 - e^{\theta}\right)^{-n}$ with $\theta < 0$, $\mu(\theta) = ne^{\theta}\left(1 - e^{\theta}\right)^{-1}$,
$H(z) = (1-z)^{-n}$ with $R_H = 1$, and $H^{(k)}(0) = c_k^*$ for all $k \in \mathbb{N}$

Construction II: uniform consistency classes

Set $\eta = e^{\theta}$ for $\theta \in \Theta$ and $\boldsymbol{\eta} = \left(e^{\theta_1}, \ldots, e^{\theta_m}\right)$.

Theorem (Uniform consistency for Construction II)

Let $\mathscr{F}$ be an NEF generated by β with support $\mathbb{N}$ and ρ a finite, positive constant. If $\frac{1}{c_k k!} \le \frac{C}{k!}$ for all $k \in \mathbb{N}$, then a uniform consistency class is

$$\mathscr{Q}_{\mathrm{II},1}\left(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma\right) = \left\{ \begin{array}{c} \|\boldsymbol{\theta}\|_{\infty} \le \rho, \pi_{1,m} \ge m^{(\gamma-1)/2}, \\ t = 2^{-1} \left\|\boldsymbol{\eta}\right\|_{\infty}^{-1/2} \gamma \log m, \\ \lim_{m \to \infty} t \min_{i \in I_{1,m}} \left|\eta_0 - \eta_i\right| = \infty \end{array} \right\}$$

for any fixed $\gamma \in (0, 1]$.

Construction II: uniform consistency classes

- The GF of Poisson family has a constant derivative sequence at 0: $H^{(k)}(0) = c_k k! = 1$ for all $k \in \mathbb{N}$

Theorem (Uniform consistency for Construction II)

For Poisson family, a uniform consistency class is

$$\mathcal{Q}_{\mathrm{II},2}(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma) = \left\{ \begin{array}{c} \|\boldsymbol{\theta}\|_\infty \leq \rho, \pi_{1,m} \geq m^{(\gamma'-1)/2}, \\ t = \sqrt{\|\boldsymbol{\eta}\|_\infty^{-1/2} \gamma \log m}, \\ \lim_{m\to\infty} t \min_{i \in I_{1,m}} \left|\eta_0 - \eta_i\right| = \infty \end{array} \right\}$$

for any fixed $\gamma \in (0,1)$ *and* $\gamma' > \gamma$.

Construction III: continuous NEFs with separable moments

- Assume $0 \in \Theta$, so that β is a probability measure with finite moments of all orders
- Let

$$\tilde{c}_n(\theta) = \frac{1}{L(\theta)} \int x^n e^{\theta x} \beta(dx) = \int x^n dG_\theta(x) \quad \text{for} \quad n \in \mathbb{N} \tag{13}$$

  be the moment sequence for $G_\theta \in \mathscr{F}$

- Fact: (13) is the Mellin transform of the measure G_θ
- Caution: Mellin transform is not Fourier transform even though they are connected

Definition (NEF with separable moments)

If there exist two functions $\zeta, \xi : \Theta \to \mathbb{R}$ and a sequence $\{\tilde{a}_n\}_{n\geq 0}$ that satisfy the following:

- $\xi(\theta) \neq \xi(\theta_0)$ whenever $\theta \neq \theta_0$, $\zeta(\theta) \neq 0$ for all $\theta \in U$, and ζ does not depend on any $n \in \mathbb{N}$,
- $\tilde{c}_n(\theta) = \xi^n(\theta)\zeta(\theta)\tilde{a}_n$ for each $n \in \mathbb{N}$ and $\theta \in \Theta$,
- $\Psi(t,\theta) = \sum_{n=0}^{\infty} \frac{t^n \xi^n(\theta)}{\tilde{a}_n n!}$ is absolutely convergent pointwise in $(t,\theta) \in \mathbb{R} \times \Theta$,

then the moment sequence $\{\tilde{c}_n(\theta)\}_{n\geq 0}$ is called "separable" (at θ_0).

Theorem (Construction for NEFs with Separable Moments)

Assume that the NEF $\mathscr{F}$ has a separable moment sequence $\{\tilde{c}_n(\theta)\}_{n\geq 0}$ at θ_0. For $t, x \in \mathbb{R}$ set

$$K\left(t,x;\mu_0\right) = \frac{1}{\zeta(\theta_0)} \int_{[-1,1]} \sum_{n=0}^{\infty} \frac{(-tsx)^n \cos\left(\frac{\pi}{2} n + ts\xi(\theta_0)\right)}{\tilde{a}_n n!} \omega(s)\, ds. \quad (14)$$

Then

$$\psi\left(t,\mu;\mu_0\right) = \int K(t,x;\theta_0)\, dG_\theta(x)$$
$$= \frac{\zeta(\theta)}{\zeta(\theta_0)} \int_{[-1,1]} \cos\left(ts(\xi(\theta_0) - \xi(\theta))\right) \omega(s)\, ds. \quad (15)$$

Construction III: one example

Let Γ be Euler's Gamma function. Gamma family $\mathsf{Gamma}(\theta, \sigma)$:

- Its basis $\frac{d\beta}{d\nu}(x) = \frac{x^{\sigma-1}e^{-x}}{\Gamma(\sigma)} 1_{(0,\infty)}(x)\, dx$ with $\sigma > 0$
- $L(\theta) = (1-\theta)^{-\sigma}$ and $\mu(\theta) = \frac{\sigma}{1-\theta}$
- CDF and PDF:

$$\frac{dG_\theta}{d\nu}(x) = f_\theta(x) = (1-\theta)^\sigma \frac{e^{\theta x} x^{\sigma-1} e^{-x}}{\Gamma(\sigma)} 1_{(0,\infty)}(x) \quad \text{for } \theta < 1$$

- The Gamma family subsumes Exponential family and central Chi-square family

Construction III: one example

- Fact:

$$\tilde{c}_n(\theta) = (1-\theta)^{\sigma} \int_0^{\infty} \frac{e^{-y} y^{n+\sigma-1}}{\Gamma(\sigma)} \frac{dy}{(1-\theta)^{n+\sigma-1}} = \frac{\Gamma(n+\sigma)}{\Gamma(\sigma)} \frac{1}{(1-\theta)^n},$$

which implies $\xi(\theta) = (1-\theta)^{-1}$, $\tilde{a}_n = \frac{\Gamma(n+\sigma)}{\Gamma(\sigma)}$ and $\zeta \equiv 1$

- Construction III:

$$K(t,x;\theta_0) = \int_{[-1,1]} \omega(s) \sum_{n=0}^{\infty} \frac{(-tsx)^n \Gamma(\sigma) \cos\left(\frac{\pi}{2} n + \frac{ts}{1-\theta_0}\right)}{n! \Gamma(\sigma+n)} ds,$$

for which

$$\psi(t,\theta;\theta_0) = \int_{[-1,1]} \cos\left(ts\left((1-\theta_0)^{-1} - (1-\theta)^{-1}\right)\right) \omega(s)\, ds$$

- Recall Construction I applied to location-shift family:

$$\begin{aligned}\psi\left(t,\mu;\mu_0\right) &= \int K\left(t,x;\mu_0\right) dF_{\mu}\left(x\right) \\ &= \int K\left(t,y+\left(\mu-\mu_0\right);\mu_0\right) dF_{\mu_0}\left(y\right),\end{aligned}$$

where Fourier transform: $K\left(t,x;\mu_0\right) \mapsto K\left(t,y+\left(\mu-\mu_0\right);\mu_0\right)$

- Construction III for Gamma family:

$$\begin{aligned}\psi\left(t,\theta;\theta_0\right) &= \int K\left(t,x;\theta_0\right) dG_{\theta}\left(x\right) \\ &= \int K\left(t,\frac{y}{1-\theta};\theta_0\right) \beta\left(dy\right),\end{aligned}$$

where Mellin transform: $K\left(t,x;\theta_0\right) \mapsto K\left(t,y\left(1-\theta\right)^{-1};\theta_0\right)$

- Set $u_{3,m} = \min_{1 \le i \le m} \{1 - \theta_i\}$ and $\xi(\theta) = (1-\theta)^{-1}$

Theorem (Uniform Consistency of Construction III)

Consider Gamma family with a fixed $\sigma > 0$. Let $\rho > 0$ be a finite constant. If $\sigma > 3/4$, then a uniform consistency class is

$$\mathcal{Q}_{\mathrm{III}}\left(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma\right) = \left\{ \begin{array}{c} \|\boldsymbol{\theta}\| \le \rho, t = 4^{-1}\gamma u_{3,m} \log m, \\ \lim_{m\to\infty} u_{3,m} \log m = \infty, \\ \pi_{1,m} \ge m^{(\gamma-1)/2}, \\ \lim_{m\to\infty} t \min_{i \in I_{1,m}} \left|\xi(\theta_0) - \xi(\theta_i)\right| = \infty \end{array} \right\}$$

for any fixed $\gamma \in (0,1]$.

Theorem (Uniform Consistency of Construction III)

Consider Gamma family with a fixed $\sigma > 0$. Let $\rho > 0$ be a finite constant. If $\sigma \leq 3/4$, then

$$\mathcal{Q}_{\mathrm{III}}\left(\boldsymbol{\theta}, t, \pi_{1,m}; \gamma\right) = \left\{ \begin{array}{c} \|\boldsymbol{\theta}\| \leq \rho, t = 4^{-1}\gamma u_{3,m} \log m, \\ \pi_{1,m} \geq m^{(\gamma'-1)/2}, \\ \lim_{m\to\infty} t \min_{i \in I_{1,m}} \left|\xi\left(\theta_0\right) - \xi\left(\theta_i\right)\right| = \infty \end{array} \right\}$$

for any fixed $\gamma \in (0,1)$ and $\gamma' > \gamma$.

Summary

- Solutions to Lebesgue-Stieltjes integral equation can serve as universal construction for proportion estimators
- Under independence, the induced estimators are often uniformly consistent
- The uniform consistency class and speed of convergence for each such estimator are largely determined by
  - the modulus of the CF corresponding to the null parameter for families with Riemann-Lebesgue type CFs (including many location-shift families)
  - the radius of convergence of the generating function of discrete natural exponential families
  - the decaying speed of the moment sequence of continuous natural exponential families

A few remaining tasks

- Construct estimators via solutions to Lebesgue-Stieltjes integral equations for $\pi_{1,m}$ for composite null hypotheses such as
  - $H_{i0} : \mu_i \in (a, b) \quad \text{vs} \quad H_{i1} : \mu_i \notin (a, b)$
  - $H_{i0} : \mu_i \leq a \quad \text{vs} \quad H_{i1} : \mu_i > a$
- Uniform consistency of such estimators under dependence outside Gaussian families and their associated mixtures
- Classify natural exponential families with separable moment functions

A few remaining tasks

- Classify distributions whose CFs satisfy

$$\left\{t \in \mathbb{R} : \hat{F}_{\mu}(t) = 0\right\} = \varnothing \quad \text{for} \quad \mu \in U' \tag{16}$$

- Completely understand the relationships between families with Riemann-Lebesgue type CFs, those whose CFs satisfy (16), and infinitely divisible distributions
- Concentration inequalities for sums of non-Lipschitz, unbounded transforms of independent randomize variables

- Kevin Vixie, Hong-Ming Yin, Charles N. Moore, Sheng-Chi Liu for discussions on solutions to integral and algebraic equations
- Gérard Letac for guidance on research involving natural exponential families
- Sergey Lapin and Ovidiu Costin for providing access to two papers
- Jiashun Jin for comments on manuscript for the work
- Mark D. Ward for comments on coefficients of exponential generating functions
- New Faculty Seed Grant by Washington State University

Key References

- Xiongzhi Chen, *Estimators of the proportion of false null hypotheses: I "universal construction via Lebesgue-Stieltjes integral equations and uniform consistency under independence"*, arxiv preprint
- References in the above preprint