Readings on Moment Inequality Methods

Manu Navjeevan

June 20, 2020

Contents

1 A Practical Method for Testing Many Moment Inequalities; *Yuehao Bai, Andres Santos, Azeem M. Shaikh*

1.1 Introduction

Setup: $\{X_i\}_{i=1}^n$ i.i.d with distribution $P \in P_n$ on R^n. Consider the problem of testing

$$H_0 : P \in \mathbf{P}_{0,n} \quad \text{versus} \quad H_1 : P \in \mathbf{P}_{1,n} \tag{1}$$

where

$$\mathbf{P}_{0,n} \equiv \{P \in \mathbf{P}_n : E_P[X_i] \leq 0\} \tag{2}$$

and $\mathbf{P}_{1,n} = \mathbf{P}_n / \mathbf{P}_{0,n}$. The inequality in 2 is interpreted component wise and $\mathbf{P}_n$ is a large class of possible distributions for the observed data. Indexing both the number of moments p_n and the class of possible distributions by the sample size allows for the number of moments to grow (rapidly) with the sample size n. Goal is to construct test that are uniformply consistent in level; i.e

$$\limsup_{n\to\infty} \sup_{P \in \mathbf{P}_{0,n}} E_P[\phi_n] \leq \alpha \tag{3}$$

A test can be viewed as a function of the data $\phi_n = \phi_n : \mathcal{X}^n \to \{0,1\}$ where $\mathcal{X}^n$ is generally some subset of $\mathbb{R}^n$ where the data takes its values.

There are a large class of problems in economics in which the number of moments is large. For example, in the entry models as in Cilberto and Tamer (2009) the number of moment inequalities to check is $p_n = o(2^{m+1})$ where m is the number of firms. Apart from Chernozukhov et. al (2019), this has typically been done by limiting $\mathbf{P}_n$ so that the number of moments p_n are small. Canay and Shaikh (2017) provide a detailed review of these tests. This paper focuses on the two step testing procedure of Romano et. al (2014). Test is shown to satisy (3) under assumptions on $\mathbf{P}_n$ that restrict p_n to not depend on n. However, the test is "practical" in that it is computationally feasible even if the number of moments is large. **Paper shows that the test of Romano et. al (2014) continues to satisy (3) for a large class of distributions that permits the number of moments p_n to grow exponentially with the sample size n.**

Theoretical analysis relies on Chernozhukov et. al (2013, 2017) on the high dimensional CLT. This is seminal work. Allen (2018) argues that the test proposed Romano et al. (2014) is more powerful in finite samples then the test proposed by Chernozhukov et al. (2019).

1.2 Main Result

Begin this section by describing the testing procedure in Romano et al. (2014). To do so, best to introduce some further notation. For $1 \leq j \leq p_n$ let $X_{i,j}$ denote the jth component of X_i and set

$$\bar{X}_{j,n} \equiv \frac{1}{n} \sum_{i=1} n X_{i,j} \tag{4}$$

$$S_{j,n}^2 \equiv \frac{1}{n} \sum_{i=1}^{n} (X_{i,j} - \bar{X}_{j,n})^2 \tag{5}$$

Can also use the notation $\mu_j(p) \equiv E_P[X_{i,j}]$ and $\sigma_j^2(P) \equiv \mathrm{Var}_P[X_i, j]$ so that (4) and (5) can be expressed as $\mu_j(\hat{P}_n)$ and $\sigma_j^2(\hat{P}_n)$, respectively, where $\hat{P}_n$ is the empirical distribution of $\{X_i\}_{i=1}^n$. Focus on a test that rejects for large values of

$$T_n \equiv \max\left\{\max_{1\leq j\leq p_n} \frac{\sqrt{n}\bar{X}_{j,n}}{S_{j,n}}, 0\right\}$$

In defining critical value, useful to introduce an i.i.d sequence of random variables with distribution $\hat{P}_n$ conditional on $\{X_i\}_{i=1}^n$, which we will denote $X_i^*, i = 1, \ldots, n$. Further define $\bar{X}_{j,n}^*$ and $(S_{j,n}^*)^2$ analagously to before, but substituting in X_i^*. Critical value for T_n is given by

$$\hat{c}_n^{(2)}(1-\alpha+\beta) \equiv \inf \mathcal{S}_n(1-\alpha+\beta) \tag{6}$$

where

$$\mathcal{S}_n(a) \equiv \left\{ c \in \mathbb{R} : \mathbb{P}\left[\max\left\{ \max_j \frac{\sqrt{n}(\bar{X}_{j,n}^* - \bar{X}_{j,n} + \hat{\mu}_{j,n})}{S_{j,n}^*}, 0 \right\} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq a \right\}$$

Here $\alpha \in (0, 0.5)$ is the nominal level of the test and $\beta \in (0, \alpha)$ and

$$\hat{\mu}_{j,n} \equiv \min\left\{ \bar{X}_{j,n} + \frac{S_{j,n}}{\sqrt{n}} \hat{c}_n^{(1)}(1-\beta), 0 \right\} \tag{7}$$

with

$$\hat{c}_n^{(1)} \equiv \inf\left\{ c \in \mathbb{R} : \mathbb{P}\left[\max_{1 \leq j \leq p_n} \frac{\sqrt{n}(\bar{X}_{j,n} - \bar{X}_{j,n}^*)}{S_{j,n}^*} \leq c \;\middle|\; \{X_i\}_{i=1}^n \right] \geq 1-\beta \right\}$$

The test is then

$$\phi_n^{\mathrm{RSW}} \equiv \mathbb{1}\left\{ T_n \geq \hat{c}_n^{(2)}(1-\alpha+\beta) \right\} \tag{8}$$

Motivating this choice of critical value it is useful to note that the test statistic T_n satisfies

$$T_n = \max\left\{ \max_j \left(\frac{\sqrt{n}(\bar{X}_{j,n} - \mu_j(P))}{S_{j,n}} + \frac{\sqrt{n}\mu_j(P)}{S_{j,n}} \right), 0 \right\} \tag{9}$$

Decomposition highlights that the main impedment in approximating the distribution of T_n is the presence of nuisance parameters $\sqrt{n}\mu_j(P)$ for $1 \leq j \leq p_n$.[1] Though these nuisance parameters cannot be consistently estimated, Romano et al (2014) observe that it may still be possible to construct a suitably valid confidence region for them.

Lemma in Appendix employs Romano insight and high dimensional CLT of Chernozhukov et al. (2017) to show that, under conditions that permit p_n to grow rapidly with the sample size n, $\sqrt{n}\mu_j(P) \leq \sqrt{n}\hat{\mu}_{j,n}$ for all $j \leq p_n$ with pr. approximately no less than $1-\beta$ whenever the null hypothesis in (1) is true. Since T_n is monotonically increasing in the nuisance parameters $\sqrt{n}\mu_j(P)$ for all $1 \leq j \leq p_n$ it follows that, viewed as a funtion of these nuisance parameters, any quantile of T_n is maximized over said confidence region by setting $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. Then, the critical value $\hat{c}_n^{(2)}(1-\alpha+\beta)$ is a bootstrap estimate of the $1-\alpha+\beta$ quantile of T_n under the "least favorable" nuisance parameter value $\sqrt{n}\mu_j(P) = \sqrt{n}\hat{\mu}_{j,n}$ for all j. The $1-\alpha-\beta$ quantile is employed instead of β to account for that, with pr. appx no greater than β, $\sqrt{n}\mu_j(P) > \sqrt{n}\hat{\mu}_{j,n}$. Analysis of test (8) hinges on following assumption:

Assumption 1. Assume (i) $\{X_i\}_{i=1}^n$ is an i.i.d sample with $X_i \in \mathbb{R}^{p_n}$ and $X_i \sim P \in \mathbf{P}_n$; (ii) $\sigma_j(P) > 0$ for all $1 \leq j \leq p_n$ and $P \in \mathbb{P}_n$; (iii) For $k = 1, 2$, there is a $M_{k,n} < \infty$ such that $E_P[|X_{i,j} - \mu_j(P)|^{2+k}] \leq \sigma_j^{2+k}(P) M_{k,n}^k$ for all $1 \leq j \leq p_n$ and $P \in \mathbf{P}_n$; (iv) There exists a $B_n < \infty$ such that $E_P\left[\max_{1 \leq j \leq p_n} \left|X_{i,j} - \mu_j(P)\right|^4\right] \leq B_n^4$ for all $P \in \mathbf{P}_n$; (v) $(M_{1,n}^2 \vee M_{2,n}^2 \vee B_n^2) \log^{3.5}(p_n n) = o(n^{(1-\delta)/2})$ for some $\delta \in (0,1)$

1(i) formalizes that $\{X_i\}_{i=1}^n$ be an i.i.d sample, while Assumption 1(ii) requires the variance of $X_{i,j}$ to be positive for all $P \in \mathbb{P}_n$ and $1 \leq j \leq p_n$.1(iii) imposes a uniform in P and j bound on the standardized moments of $X_{i,j}$. Condition is a strengthening of the uniform integrability requirements of Romano et al (2014) required so study a setting in which p_n diverges to infinity. Part (iv) bounds the 4th moments of the maximim of $X_{i,j}$. Finally, (v) states the main conditiion governing how fast p_n can grow with n. Under suitable moment restrictions on $X_{i,j}$, p_n may grow exponentially with n. Now ready for main result

[1] I'm not entirely sure why they cannot be consistently estimated. I think this is because we are only partially identified.

Theorem 1. *If Assumption 1 holds,* $\alpha \in (0, \frac{1}{2})$ *and* $0 < \beta < \alpha$*, then* ϕ_n^{RSW} *as defined in (8) satisfies uniform consistency in level as defined in (3)*

The rest of this paper goes through some simulations. It is also just a working paper at the moment. Probably it is best to go through the main proof; but I will print it out and make some notes on this.

2 Inference on Causal and Structural Parameters Using Many Moment Inequalities; *Victor Chernozhukov, Denis Chetverikov, and Kengo Kato (ReStud, 2019)*

2.1 Introduction

In recent years, moment inequalities framework has developed into a powerful tool for inference on causal and structural parameters in partially identified models. Many papers study models with a finite and fixed number of conditional and unconditional moment inequalities. IN practice the number of moment inequalities implied by the model is often large.

Examples of testing (very) many moment inequalities

- Consumer is selecting a bundle of products for purchase and moment inequalities come from revealed preference argument (Pakes, 2010)
- Market structure model of Ciliberto and Tamer (2009), number of moment inequalities equals the number of possible combinations of firms that could potentially enter the market (grows exponentially in the number of firms)
- Dynamic model of imperfect competition of Bajari, Benkard, Levin (2007)m where deviations from optimal policy serve to define many moment inequalities
- Beresteanu, Molchanov, Molinari (2011), Galichon and Henry (2011)[1], Chesher, Rosen, Smolinksi (2013), and Chester and Rosen (2013)

Many examples have inportant in that the many inequalities under consideration are "unstructured", they cannot be viewed as unconditional moment inequalities generated froma small number of conditional inequalities with a low-dimensional conditioning variable. So existing inference methods for conditional moment inequalities, though fruitful in many cases

Formally describing the problem, let $\{X_i\}_{i=1}^n$ be a sequence of i.i.d random vectors in $\mathbb{R}^p$, where $X_i = (X_{i1}, \dots, X_{ip})^T$, with a common distribution denoted by $\mathcal{L}_X$. For $j \leq p$, we write $\mu_j := \mathbb{E}[X_{1j}]$. Interested in testing the null hypothesis

$$H_0 : \mu_j \leq 0 \text{ for all } j = 1, \dots, p \tag{1}$$

Against the alternative

$$H_1 : \mu_j > 0 \text{ for some } j = 1, \dots, p \tag{2}$$

Refer to (1) as the moment inequalities and say the jth moment is satisfied (viloated) if $\mu_j \leq 0$ ($\mu_j > 0$). Paper will allow number of moment inequalities $p \gg n$. Consider a test statistic given by the maximum over p Studentized (t-type) inequality specific statistic. Consider critical values based upon (i) the union bound combined with a moderate deviation inequality for self-normalized sums and (ii) bootstrap methods. Among bootstrap methods, consider multiplier and emppirical bootstrap methods. These are simulation based and computationally more difficult, but take into account correlation structure and yeild lower critical values. SN method is particularly useful for grid search when the researcher is interested in constricting a confidence interval for identified set.

Also consider two-step methods incorporating inequality selection procedutes. Two-step methods get rid of most uninformative inequalities, that is inequalities with $\mu_j < 0$ if μ_j is not too close to 0. Also develop novel three-step methods by incorporating double inequality selection procedures. These are suitable in parametric modesl defined via momoent inequalities and allow to drop weakly informative inequalities in addition to uninformative inequalities.[2]. Results can be used for construction of confidence regions for identifiable parameters in partially identified models defined by moment inequalities. Show that results are asymptotically honest (don't quite know what this means).

Literature testing unconditional moment inequalities is large. See White (2000), Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and

[1] This seems like a good place to start reading

[2] Can be extended to nonparametric models as well

Soares (2010), Canay (2010), Bugni (2011), Andrews and Jia-Barwick (2012), and Romano, Shaikh, and Wolf (2014).

In this paper we implicitly assume that $X_1, \dots, X_n$ and p are indexed by n. Mainly interested in the case that $p = p_n \to \infty$ as $n \to \infty$

2.2 Motivating Examples

Section provides examples that motivate the framework where the number of moment inequalities p is large and potentially much larger than the sample size n. In these examples, one actually has many conditional rather than unconditional inequalities. Results cover conditioning as well.

2.2.1 Market Structure Model

Let m denote the number of firms that could potentially enter the market. Let m-tuple $D = (D_1, \dots, D_m)$ denote entry decisions of these firms. That is, $D_j = 1$ if the firm j enters the market and $D_j = 0$ otherwise. Let $\mathcal{D}$ denote the possible values of D. We have that $|\mathcal{D}| = 2^m$.

Let X and ϵ denote the (exogeneous) characteristics of the market as well as characteristics of the firms that are observed and not observed by the researcher, respectively. The profit of the firm j is given by

$$\pi_j(D, X, \epsilon, \theta)$$

where π_j is known up to a parameter θ. Both X and ϵ are observed by the firms and a Nash Equilibrium is played so that, for each j,

$$\pi_j((D_j, D_{-j}), X, \epsilon, \theta) \geq \pi_j((1 - D_j, D_{-j}), X, \epsilon, \theta)$$

D_{-j} denotes the decisions of all firms excluding the firm j. Then one can find set-valued functions $R_1(d, X, \theta)$ and $R_2(d, X, \theta)$ such that d is the unique equilibrium whenever $\epsilon \in D_1(d, X, \theta)$ and d is *an* equilibrium whenever $\epsilon \in R_2(d, X, \theta)$. In the second case, the probability that the researcher sees d as an equilibrium depends on the equilibrium selection mechanism. Without further information, anything can be in $[0, 1]$. Therefore we have the following bounds

$$\begin{aligned} \mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta\}|X] &\leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \\ &\leq \mathbb{E}[\mathbb{1}\{\epsilon \in R_1(d, X, \theta) \cup R_2(d, X, \theta)\}|X] \end{aligned}$$

Further assuming that the conditional distribution of ϵ given X is known (or known up to a parameter that is part of θ), both the LHS and RHS of these inequalites can be calculated. Denote them $P_1(d, X, \theta)$ and $P_2(d, X, \theta)$, respectively to obtain

$$P_1(d, X, \theta) \leq \mathbb{E}[\mathbb{1}\{D = d\}|X] \leq P_2(d, X, \theta) \tag{3}$$

for all $d \in \mathbb{D}$. These can be used for inference on the parameter θ. Note that the number of inequalities in (3) is $2|\mathcal{D}| = 2^{m+1}$. This is a large number, even if m is moderately large. Moreover, these inequalities are conditional on X. So, they can be transformed into a large and increasing number of unconditional moment inequalities as described above. Also, if the firms have more than two decisions, the number of inequalities will be even larger.

Some other examples are given, but I won't cover them in notes.

2.3 Test Statistic

Begin preparing some notation. Assume that

$$\mathbb{E}[X_{1,j}^2] < \infty, \sigma_j^2 := \mathrm{Var}(X_{1,j}) > 0, j = 1, \dots, p \tag{4}$$

For $j = 1, \dots, p$ let $\hat{\mu}_j$ and $\hat{\sigma}_j$ be the sample mean and variance of $\{X_{i,j}\}_{i=1}^n$. Many different possible test statistics. Somewhat natural to consider statistics that take large values thehn some of $\hat{\mu}_j$'s are large. In this paper focus on statistic that takes large values when at least one of $\hat{\mu}_j$ are large.

In specific, focus on the following test statistic:

$$T = \max_{1 \le j \le p} \frac{\sqrt{n}\,\hat{\mu}_j}{\hat{\sigma}_j} \tag{5}$$

Large values of T indicate a likely violation of H_0, so it is natural to consider tests of the form

$$T > c \implies \text{ reject } H_0$$

where c is appropriately chosen so that the test approximately has size $\alpha \in (0, 1)$. Consider various ways for calculating critical values and prove their validity.

2.4 Critical Values

Now move to define critical values for T such that under H_0, the probability of rejecting H_0 does not exceed size α asymptotically. Methods are ordered bu increasing computational complexity, increasing strength of required conditions, and also increasing power. Basic idea for the construction of critical values for T lies in the fact, that, under H_0:

$$T \le \max_{1 \le j \le p} \frac{\sqrt{n}(\hat{\mu}_j - \mu_j)}{\hat{\sigma}_j}$$

Consider two approaches to constructing such critical values: self-normalized and bootstrap methods. Also consider two- and three-step variants of the methods by incorporating inequality selection.

Following notation used:

$$Z_{ij} = (X_{ij} - \mu_j)/\sigma_j \quad \text{and} \quad Z_i = (Z_{i1}, \dots, Z_{ip})^T$$

Observe that $\mathbb{E}[Z_{ij}] = 0$ and $\mathbb{E}[Z_{ij}^2] = 1$. Define

$$M_{n,k} = \max_{1 \le j \le p} \left(\mathbb{E}\left[|Z_{1,j}|^k \right] \right)^{1/k}, k = 3, 4, \quad \text{and } B_n = \left(\mathbb{E}\left[\max_{1 \le j \le p} Z_{1j}^4 \right] \right)^{1/4}$$

The dependence on n comes via the dependence of $p = p_n$ on n implicitly. By Jensen's inequality, $B_n \ge M_{n,4} \ge M_{n,3} \ge 1$. In addition, if all Z_{ij}'s are bounded a.s by a constant C, we have that $C \ge B_n$. These are useful to get a sense of various conditions on $M_{n,3}, M_{n,4}$ and B_n imposed in the theorems below.

2.4.1 Self Normalized Critical Values

One-step method: Self-normalized method considered is based on the union bound combined with moderate deviation inequality for self-normalized sums. Under H_0

$$\mathbb{P}(T > c) \le \sum_{j=1}^{p} \mathbb{P}(\sqrt{n}(\hat{\mu}_j - \mu_j)/\,\hat{\sigma}_j > c) \tag{6}$$

This bound seems crude when p is large. However, will exploit the self normalizing $\sqrt{n}(\hat{\mu}_j - \mu_j)/\,\hat{\sigma}_j$ to show that RHS of above is bounded, even if c is growing logarithmically fast with p. Using such a c will yield a test with better power properties.

For $j = 1, \dots, p$, define

$$U_j := \sqrt{n}\mathbb{E}_n[Z_{ij}]/\sqrt{\mathbb{E}_n[Z_{ij}^2]}$$

Simple algebra yields, we see that

$$\sqrt{n}(\hat{\mu}_j - \mu_j)/\hat{\sigma}_j = U_j/\sqrt{1 - U_j^2/n}$$

where the right-hand side is increasing in U_j as long as $U_j \geq 0$. So under H_0,

$$\mathbb{P}(T > c) \leq \sum_{j=1}^{p} \mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right), \ c \geq 0 \tag{7}$$

Moderate deviation inequality for self-normalized sums of Jing, Shao, and Wang (2003) implies that for moderatley large $c \geq 0$,

$$\mathbb{P}\left(U_j > c/\sqrt{1 + c^2/n}\right) \approx \mathbb{P}\left(Z > x/\sqrt{1 + c^2/n}\right)$$

where $Z \sim N(0, 1)$. The above approximation holds even if Z_{ij} only have $2 + \delta$ finite moments for some $\delta > 0$. Therefore, take the critical value as

$$c^{SN}(\alpha) = \frac{\Phi^{-1}(1 - \alpha/p)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/p)^2/n}} \tag{8}$$

where $\Phi(\cdot)$ is the normal cdf. We call $c^{SN}(\alpha)$ the one-step SN critical value with size α as it's derivation depends on the moderate deviation inequality for self-normalized sums. Note that

$$\Phi^{-1}(1 - \alpha/p) \sim \sqrt{\log(p/\alpha)}$$

so $c^{SN}(\alpha)$ depends on p only through $\log(p)$. Following theorem provides a non asymptotic bound on the probability that the rest statsistic T exceeds the SN critical value $c^{SN}(\alpha)$ under H_0 and shows that the bound converged to α under mild regularity conditions, validating the SN method.

Theorem 1. *(Validity of one-step SN method). Suppose that $M_{n,3}\Phi^{-1}(1 - \alpha/p) \leq n^{1/6}$. Then under H_0,*

$$\mathbb{P}(T > c^{SN}(\alpha)) \leq \alpha \left[1 + Kn^{-1/2} M_{n,3}^3 \left\{1 + \Phi^{-1}(1 - \alpha/p)\right\}^3\right]$$

where K is a universal constant. Hence, if there exists constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that

$$M_{n,3}^3 \log^{3/2}(p/\alpha) \leq C_1 n^{1/2 - c_1} \tag{9}$$

then there exists a positive constant C depending only on C_1 such that under H_0,

$$\mathbb{P}(T > c^{SN}(\alpha)) \leq \alpha + Cn_{-c_1} \tag{10}$$

Moreover, this bound holds uniformly over all distributions $\mathcal{L}_{\mathfrak{X}}$ satisfying the moment conditions as well as the above requirement (9). In addition, if (9) holds, all components of X_1 are independent, $\mu_j = 0$ for all $1 \leq j \leq p$ and $p = p_n \to \infty$, then

$$\mathbb{P}(T > c^{SN}(\alpha)) \to 1 - e^{-\alpha}$$

I think the last bit is just to show that the test is approximately non-conservative.

Two-step method: Now move to combine the SN method with inequality selection. Motivation for doing this is that when $\mu_j < 0$ for some $j = 1, \ldots, p$ the inequality in (6) becomes strict. So, when there are many j for which μ_j are negative and large in absolute value, the resulting rest with one-step SN critical values would tend to be unnecessarily conservative. So, in order to improve the power of the test, it is better to exclude j for which μ_j are below some (negative) threshold when computing critical values.

Formally, let $0 < \beta_n < \alpha/2$ be some constant. For generality, allow β_n to depend on n. In particular, we allow $\beta_n = o(1)$. Let $c^{SN}(\beta_n)$ be the SN critical value with size β_n and define the set $\hat{J}_{SN} \subset \{1, \dots, p\}$ by

$$\hat{J}_{SN} := \left\{ j \in \{1, \dots, p\} : \sqrt{n}\hat{\mu}_j / \hat{\sigma}_j > -2c^{SN}(\beta_n) \right\} \tag{11}$$

Let $\hat{k} = |\hat{J}_{SN}|$. Then, the two step SN critical value is defined by

$$c^{SN,2S}(\alpha) = \begin{cases} \frac{\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}{\sqrt{1-\Phi^{-1}(1-(\alpha-2\beta_n)/\hat{k})}}, & \text{if } \hat{k} \geq 1 \\ 0, & \text{if } \hat{k} = 0 \end{cases} \tag{12}$$

Then paper claims the following theorem

Theorem 2. *Suppose there exist constants* $0 < c_1 < 1/2$ *and* $C_1 > 0$ *such that*

$$M_{n,3}^3 \log^{3/2}\left(\frac{p}{\beta_n \wedge (\alpha - 2\beta_n)}\right) \leq C_1 n^{1/2-c_1}$$
$$\text{and } B_n^2 \log^2(p/\beta_n) \leq C_1 n^{1/2-c_1}$$

Then there exist positive constants c, C *depending only on* α, c_1, C_1 *such that under* H_0,

$$\mathbb{P}(T > c^{SN,2S}(\alpha)) \leq \alpha + Cn^{-c} \tag{13}$$

Moreover, this bound holds uniformly over all distribution $\mathcal{L}_X$ *satisfying (6) and the above condition. In addition, if all components of* X_1 *are independent,* $\mu_j = 0$ *and* $p = p_n \to \infty$ *while* $\beta_n \to 0$ *then*

$$\mathbb{P}(T > c^{SN,2S}(\alpha)) \to 1 - e^{-\alpha}$$

2.4.2 Bootstrap Methods

Section considers Multiplier Bootstrap and Empitical Bootstrap methods. These methods are computationally harder but they lead to less conservative tests.

One-Step Method First consider the one-step method (without moment selection). In order to make the test have size α, it is enough to choose the critical value as a bound on the $(1-\alpha)$ quantile of the distribution of

$$\max_{1 \leq j \leq p} \sqrt{n}(\hat{\mu} - \mu_j)/\hat{\sigma}_j$$

The self normalizing method finds such a bound using the union bound and moderate deviation inequality for self-normalized sums. However, SN method may be conservative as it ignores correlation between the coordinates in X_i.

Alternatively, we consider a Gaussian approximation. Under suitable regularity conditions

$$\max_{i \leq j \leq p} \sqrt{n}(\hat{\mu} - \mu_j)/\hat{\sigma}_j \approx \max_{1 \leq j \leq p} \sqrt{n}(\hat{\mu}_j - \mu_j)/\sigma_j = \max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}]$$

where $Z_i = (Z_{i1}, \dots, Z_{ip})^T$ are defined above ($Z_j = (X_j - \mu_j)/\sigma_j$). When p is fixed, the central limit theorem shows that, as $n \to \infty$,

$$\sqrt{n}\mathbb{E}_n[Z_i] \rightsquigarrow Y, \text{ with } Y = (Y_1, \dots, Y_p)^Y \sim N(0, \mathrm{E}[Z_1 Z_1^T])$$

By the continuous mapping theorem, this gives us that

$$\max_{1 \leq j \leq p} \sqrt{n}\mathbb{E}_n[Z_{ij}] \rightsquigarrow \max_{1 \leq j \leq p} Y_j$$

so we can take the critical value to be the $(1-\alpha)$ wuantile of $\max_{1 \leq j \leq p} Y_j$. This theory does not cover when

p grows with n. Different tools should be used to derive an appropriate critical value for the test. A possible approach is to use a Berry-Esseen theorem that provides a suitable non-asymptotic bound between the distributions of $\sqrt{n}\mathbb{E}_n[Z_i]$ and Y. However, such Berry Esseen bounds require p to be small in comparasion with n in order to garuntee that the distribution of $\sqrt{n}\mathbb{E}_n Z_i$ is similar to that of Y. This approach builds on the work of (Chernozhukov, Chetverikov, and Kato, 2013, 2017) to show that, under some mild regularity conditions, the distribution of $\max_{1\le j\le p}\sqrt{n}\mathbb{E}_n[Z_{ij}]$ can be approximated by that of $\max_{1\le j\le p}$ in the sense of Kolmogrov distance even when p is larger or much larger than n.

Still, the distribution of $\max_{1\le j\le p} Y_j$ is typically unknown because the covariance structure of Y is unknown. So we will approximate the distribution of $\max_{1\le j\le p} Y_j$ by one of the following two bootstrap procedures:

Algorithm (Multiplier bootstrap)

1. Generate independent standard normal variables $\epsilon_1, \dots, \epsilon_n$ independent of the data
2. Construct the multiplier bootstrap test statistic

$$W^{MB} = \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_i(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} \tag{14}$$

3. Calculate $c^{MB}(\alpha)$ as the conditional $(1-\alpha)$-quantile of W^{MB} given X_1^n

Algorithm (Empirical bootstrap)

1. Generate a bootstrap sample $X_1^*, \dots X_n^*$
2. Construct the empirical bootstrap test statistic

$$W^{EB} = \max_{1\le j\le p} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^* - \hat{\mu}_j]}{\hat{\sigma}_j} \tag{15}$$

3. Calculate $c^{EB}(\alpha)$ as the contional $(1-\alpha)$ quantile of W^{EB} given X_1^n.

We call these the one step multipler bootstrap and empirical bootsrap critical values, respectively, with size α. Can be computed with any precision using simulation.

Intuitively it is expected that the multiplier bootstrap works well since, conditional on the data, the vector

$$\left(\frac{\sqrt{n}\mathbb{E}[\epsilon_i(x_{ij} - \hat{\mu}_j)]}{\sigma_j}\right)_{1\le j\le p}$$

has the centered normal distribution with covariance matrix

$$\mathbb{E}_n\left[\frac{(X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j}\frac{(X_{ik} - \hat{\mu}_k)}{\hat{\sigma}_k}\right], 1 \le j,k \le p \tag{16}$$

which should be close to the covariance matrix of the vector Y. Indeed by Theorem 2 in Chenozhukov, Chetverikov, and Kato (2015), the primary factor for the bound on the Kolmogorov [3] distance between the conditional distribution of W and the distribution of $\max_{1\le j\le p} Y_j$ is

$$\max_{1\le j,k\le p}\left|\mathbb{E}_n\left[\frac{(X_{ij} - \hat{\mu}_j)}{\hat{\sigma}_j}\frac{(X_{ik} - \hat{\mu}_k)}{\hat{\sigma}_k}\right] - \mathbb{E}[Z_{1j}Z_{1k}]\right|$$

which is shown to be small even when $p \gg n$ (under suitable conditions).

Following theorem establishes validity of the MB and EB critical values.

[3] The Kolmogorov Distance is defined as, for two pr. measures μ, ν on $\mathbb{R}$, $\text{Kolm}(\mu,\nu) := \sup_{x\in\mathbb{R}}\left|\mu\left((-\infty,x]\right) - \nu\left((-\infty,x]\right)\right|$

Theorem 3 (Validity of one-step MB and EB methods)**.** *Let $c^B(\alpha)$ stand for either $c^{MB}(\alpha)$ or $c^{EB}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn) \leq C_1 n^{1/2-c_1} \tag{17}$$

Then there exist positive constants c, C depending only on c_1, C_1 such that, under H_0,

$$\mathbb{P}(T < c^B(\alpha)) \leq \alpha + Cn^{-c} \tag{18}$$

In addition, if $\mu_j = 0$ for all j, then

$$\left|\mathbb{P}(T > c^B(\alpha)) - \alpha\right| \leq Cn^{-c} \tag{19}$$

Moreover both bounds hold uniformly over all distributions L_X satisfying the conditions (4) and (17).

Leave analysis of more general exchangeable weighted bootstraps in the high dimensional setting for future works. Also observe that the condition (17) required fof the validity of the one-step MB/EB methods is stronger than qhat is required for validity of the two-step SN method.

Two-step Methods Now consider combining bootstrap methods with inequality selection. To describe, let $0 < \beta_n < \alpha/2$ be some constant. As before, β_n can depend on n. Let $c^{MB}(\beta_n)$ and $c^{EB}(\beta_n)$ be one-step MB and EB critical values with size β_n, repsectively. Define the sets $\hat{J}_{MB}$ and $\hat{J}_{EB}$ by

$$\hat{J}_B := \left\{ j \in \{1, \ldots, p\} : \sqrt{n}\,\hat{\mu}_j \,/\, \hat{\sigma}_j > -2c^B(\beta_n) \right\}$$

Then, the two-step MB and EB critical values $c^{MB,2S}(\alpha)$ and $c^{EB,2S}(\alpha)$ are defined by the following procedures

Algorithm (Multiplier bootstrap with inequality selection).

1. Generate independent standard normal random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data X_1^n.
2. Construct the multiplier bootstrap test statistic

$$W_{\hat{J}_{MB}} = \begin{cases} \max_{j \in \hat{J}_{MB}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_n(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} & \text{if } \hat{J}_{MB} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calculate $c^{MB,2S}$ as the conditional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{MB}}$ given the data

Algorithm (Empirical bootstrap with inequality selection).

1. Generate a bootstrap sample $X_1^*, \ldots, X_n^*$ as i.i.d draws from the empirical distribution of $X_1^n = \{X_1, \ldots, X_n\}$.
2. Construct the empirical bootstrap test statistic

$$W_{\hat{J}_{EB}} = \begin{cases} \max_{j \in \hat{J}_{EB}} \frac{\sqrt{n}\mathbb{E}_n[X_{ij}^* - \hat{\mu}_j]}{\hat{\sigma}_j} & \text{if } \hat{J}_{EB} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calculate $c^{EB,2S}(\alpha)$ as the conditional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{EB}}$ given the data

Theorem 4 (Validity of two-step MB and EB methods)**.** *Let $c^{B,2S}(\alpha)$ stand for either $c^{MB,2S}(\alpha)$ or $c^{EB,2S}(\alpha)$. Suppose that the asusmption of Theorem* 3 *is satisfied. Moreover, suppose that* $\log(1/\beta_n) \leq C_1 \log n$. *Then there exist positive constants c, C depending only on c_1, C_1 such that under H_0,*

$$\mathbb{P}(T > c^{B,2S}(\alpha)) \leq \alpha + Cn^{-c}$$

In addition, if $\mu_j = 0$ for all $1 \leq j \leq p$, then

$$\mathbb{P}(T > c^{B,2S}(\alpha)) \geq \alpha - 3\beta_n - Cn^{-c}$$

so that under an extra assumption that $\beta \leq C_1 n^{-c_1}$[a]

$$\left|\mathbb{P}(T > c^{B,2S}(\alpha)) - \alpha\right| \leq Cn^{-c}$$

Moreover all these bounds hold uniformly over all distributions L_X satisfying (4) and (17)

[a]which is to say β_n goes to 0 reasonable fast

It is sort of interesting to note that all these theorems are "non-asymptotic" in the sense that if the conditions hold then these inequalities "really" hold.

2.4.3 Hybrid Methods

Have considered one-step SN, MB, and EB methods and their two-step variatns. In fact, can also consider hybrids of these methods. For example, can use the SN method for inquality selection and then apply the MB or EB method for the selected inequalities, which is computationally more tractable. Notate this as the HB method. Formally, let $0 < \beta_n < \alpha/2$ be some constants and recall the set $\hat{J}_{SN} \subset \{1, \ldots, p\}$ defined above. Then the hybrid MB critical value, $c^{MB,H}(\alpha)$ is defined by the following procudure:

Algorithm (Multiplier Bootstrap Hybrid method).

1. Generate independent standard normal random variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data X_1^n.
2. Construct the bootstrap test statistic:

$$W_{\hat{J}_{SN}} = \begin{cases} \max_{j \in \hat{J}_{SN}} \frac{\sqrt{n}\mathbb{E}_n[\epsilon_n(X_{ij} - \hat{\mu}_j)]}{\hat{\sigma}_j} & \text{if } \hat{J}_{SN} \text{ is not empty} \\ 0 & \text{otherwise} \end{cases}$$

3. Calvculate $c^{MB,H}(\alpha)$ as the contional $(1 - \alpha + 2\beta_n)$-quantile of $W_{\hat{J}_{SN}}$ given the data.

This can be equibalently defined for the empirical bootstrap.

Theorem 5 (Validity of hybrid two-step methods)**.** *Let $c^{MB,H}$ stand either for $c^{MB,H}(\alpha)$ or $c^{MB,H}(\alpha)$. Suppose that there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that (17) is verified. Moreover, suppose that* $\log(1/\beta_n) \leq C_1 \log n$. *Then all the conclusions of Theorem 4 hold wilth $c^{B,MS}(\alpha)$ replaced by $c^{B,H}(\alpha)$.*

2.4.4 Three-step method

In empirical studies based on moment inqualites one generally has inequalities of the form

$$\mathrm{E}[g_j(\xi, \theta)] \leq 0 \qquad \text{for all } j = 1, \ldots, p \tag{20}$$

where ξ is a vector of r.v's from a distribution denoted $\mathcal{L}_\xi$, $\theta = (\theta_1, \ldots, \theta_r)^T$ is a vector of parameters in $\mathbb{R}^r$ and $g_1, \ldots, g_p$ a set of (known) functions. In these studies, inequalities (1) and (2) arise when one tests the null hypothesis $\theta = \theta_0$ against the alternative $\theta \neq \theta_0$ on the i.i.d data $\xi_1, \ldots \xi_n$ by setting $X_{ij} := g_j(\xi_i, \theta_0)$ and $\mu_j := \mathrm{E}[X_{1j}]$. So far, have shown how to increase the poser of such tests by employing inequality selection procedures that allow the researcher to drop uninformative inequalities. In this subsection, combine this selection procedure with another procedure suitable for the model (20) by dropping *weakly informative* inequalities, that is inequalities j with the function $\theta \mapsto \mathrm{E}[g_j(\xi, \theta)]$ being flat or nearly flat around $\theta = \theta_0$.

When the tested value θ_0 is close to some θ satisfying (20), such inequalities can only provide a weak signal of violation of the hypothesis $\theta = \theta_0$ in the sense that they have $\mu_j \approx 0$ and so it is useful to drop them. For brevity, only consider weakly informative inequality selection based on the MB and EB methods and note that similar results can be obtained for the self-normalized method. Also only consider the case where the function $\theta \mapsto g_j(\xi, \theta)$ are almost everywhere continuously differentiable and leave the extension to non-differentiable functions to future work.

Start with the necessary notation. Let $\xi_1, \ldots, \xi_n$ be a sample of observations from the distribution of ξ. Suppose that we are interested in testing the null hypothesis and alternative hypothesis

$$\begin{aligned} H_0 &: \mathrm{E}[g_j(\xi, \theta_0)] \leq 0 \quad \text{for all } j = 1, \ldots, p \\ H_a &: \mathrm{E}[g_j(\xi, \theta_0)] > 0 \quad \text{for some } j = 1, \ldots, p \end{aligned}$$

where θ_0 is some value of the parameter θ. Define

$$\begin{aligned} m_j(\xi, \theta) &:= \big(m_{j1}(\xi, \theta), \ldots, m_{jr}(\xi, \theta)\big)^T \\ &:= \left(\frac{\partial g_j(\xi, \theta)}{\partial \theta_1}, \ldots, \frac{\partial g_j(\xi, \theta)}{\partial \theta_r}\right)^T \end{aligned}$$

Further, let $X_{ij} := g_j(\xi_i, \theta_0), \mu_j := \mathrm{E}[X_{1j}], \sigma_j := \big(\mathrm{Var}(X_{1j})\big)^{1/2}, V_{ijl} := m_j l(\xi_i, \theta_0), \mu^B_{jl} = \mathrm{E}[V_{1jl}]$, and $\sigma^V_{jl} := \big(\mathrm{Var}(V_{1jl})\big)^{1/2}$. Assume that

$$\mathrm{E}[X^2_{1,j}] < \infty, \sigma_j > 0, j = 1, \ldots, p \tag{21}$$
$$\mathrm{E}[X^2_{1,j,l}] < \infty, \sigma^V_{jl} > 0, j = 1, \ldots, p, l = 1, \ldots, r \tag{22}$$

In addition, let $\hat{\mu}_j = \mathbb{E}_n[X_{ij}]$ and $\hat{\sigma}_j = \left(\mathbb{E}\left[(X_{ij} - \hat{\mu}_j)^2\right]\right)^{1/2}$ be estimators of μ_j and σ_j, respectively. Similarly let $\hat{\mu}^V_{jl} = \mathbb{E}_n[V_{ijl}]$ and $\hat{\sigma}^V_{jl} = \left(\mathbb{E}\left[(V_{ij} - \hat{\mu}^V_{jl})^2\right]\right)^{1/2}$ be esttimators of μ^V_{jl}. The inequality selection derived is similar to the bootstrap methods described in Section 4

Algorithm(Multiplier bootstrap for gradient statistic).

1. Generate independent standard normal variables $\epsilon_1, \ldots, \epsilon_n$ independent of the data.
2. Construct the multiplier bootstrap gradient statistic

$$W^V_{MB} = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[\epsilon_i(V_{ijl} - \hat{\mu}^V_{jl})]|}{\hat{\sigma}^V_{jl}} \tag{23}$$

3. For $\gamma \in (0, 1)$, calculate $c^{MB,V}(\gamma)$ as the conditional $(1 - \gamma)$ quantile of W^V_{MB} given the data.

Algorithm(Empirical bootstrap for gradient statistic).

1. Generate a bootstrap sample $V^*_1, \ldots, V^*_n$ as i.i.d draws from the data
2. Construct the empirical bootstrap gradient statistic

$$W^V_{EB} = \max_{j,l} \frac{\sqrt{n}|\mathbb{E}_n[V^*_{ijl} - \hat{\mu}^V_{jl}]|}{\hat{\sigma}^V_{jl}} \tag{24}$$

3. For $\gamma \in (0,1)$, calculate $c^{EB,V}(\gamma)$ as the conditional $(1-\gamma)$ quantile of W_{EB}^V given the data.

For $c_2, C_2 > 0$, let φ_n be a sequence of constants satisfying $\varphi_n \log n \geq c_2$ and let β_n be a sequence of constants satisfying $0 < \beta_n < \alpha/4$ and $\log\big(1/(\beta_n - \varphi_n)\big) \leq C_2 \log n$ where α is the nominal level of the test. Define three estimated sets of inequalities

$$\hat{J}_B := \Big\{ j \in \{1,\dots,p\} : \sqrt{n}\,\hat{\mu}_j \,/\, \hat{\sigma}_j > -2c^B(\beta_n) \Big\}$$
$$\hat{J}'_B := \Big\{ j \in \{1,\dots,p\} : \sqrt{n}|\,\hat{\mu}_{jl}^V \,/\, \hat{\sigma}_{jl}^V\,| > 3c^{B,V}(\beta_n - \phi_n) \text{ for some } l = 1,\dots,r \Big\}$$
$$\hat{J}''_B := \Big\{ j \in \{1,\dots,p\} : \sqrt{n}|\,\hat{\mu}_{jl}^V \,/\, \hat{\sigma}_{jl}^V\,| > c^{B,V}(\beta_n + \phi_n) \text{ for some } l = 1,\dots,r \Big\}$$

where B stands for either MB or EB.

The derived weakly informative inequality selection procedure requires that both the test statistic and the critical value depend on the estimated sets of inequalities. Let T^B and $c^{B,3S}$ denote the test statistic and the critical value. If the set $\hat{J}'_B$ is empty, set the test statistic and critical value $T^B = c^{B,3S} = 0$. Otherwise, define the test statistic

$$T^B = \max_{j \in \hat{J}'_B} \frac{\sqrt{n}\,\hat{\mu}_j}{\hat{\sigma}_j}$$

and define the three-step MB/EB critical values $c^{B,3S}(\alpha)$ for the test by the same bootstrap procedures as those for $c^{B,2S}(\alpha)$ with $\hat{J}_B$ replaced by $\hat{J}' \cap \hat{J}''_B$ and also $2\beta_n$ replaced by $4\beta_n$. That is $c^{B,3S}(\alpha)$ is the conditional $(1-\alpha+4\beta_n)$-quantile of $W_{\hat{J}'_B \cap \hat{J}''_B}$ given the data.

Stating the main results of this section requires the following notation. Let

$$Z_{ijl}^V := (V_{ijl} - \mu_{ijl}^V)/\sigma_{jl}^V \quad \text{and} \quad M_{n,k}^V := \max_{j,l} \Big(\mathrm{E}[|Z_{ijl}^V|^k]\Big)^{1/k} \quad \text{and} \quad B_n^V := \Big(\mathrm{E}[\max_{j,l}(Z_{ijl}^V)^4\Big)^{1/4}$$

Theorem 6 (Validity of three-step MB and EB methods). [a] *Let T^B and $c^{B,3S}(\alpha)$ stand for T^{MB} and $c^{MB,3S}(\alpha)$ or for T^{EB} and $c^{EB,3S}(\alpha)$. Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$\Big(M_{n,3}^3 \vee M_{n,4}^2 \vee B_n)^2 \log^{7/2}(pn)\Big) \leq C_1 n^{1/2 - c_1} \tag{25}$$

and

$$\Big((M_{n,3}^V)^3 \vee (M_{n,4}^V)^2 \vee (B_n^V)^2 \log^{7/2}(pn)\Big) \leq C_1 n^{1/2 - c_1} \tag{26}$$

Moreover, suppose that $\log(1/(\beta_n - \phi_n)) \leq C_2 \log n$ and $\phi_n \log n \geq c_2$ for some constants $c_2, C_2 > 0$. Then there exist positive constants c, C depending only on c_1, C_1, c_2, C_2such that, under H_0,

$$\mathbb{P}(T^B > c^{B,3S}(\alpha)) \leq \alpha + Cn^{-c}$$

and this bound holds uniformly over all distributions L_ξ satisfying (21),(22),(25), and (26).

[a] If I understand correctly, this method only selects out weakly uninformative inequalites based on the gradient, not necessarily inequalities with say, $\hat{\mu}_j \ll 0$.

2.5 Power

Consider the same general setup as described in the intruction and assume that (4) holds. Pick any $\alpha \in (0, 1/2)$ and consider the test of the form

$$T > \hat{c}(\alpha) \implies \text{reject } H_0$$

Where $\hat{c}(\alpha)$ is equal to $c^{SN}(\alpha), c^{SN,2S}(\alpha), c^{MB}(\alpha), c^{MB,2S}(\alpha), c^{EB}(\alpha), c^{EB,2S}(\alpha), c^{MB,H}(\alpha)$, or $c^{EB,H}(\alpha)$.[4] The following result holds:

Theorem 7 (Rate of uniform consistency)**.** *Suppose there exist constants $0 < c_1 < 1/2$ and $C_1 > 0$ such that*

$$M_{n,4}^2 \log^{1/2} p \leq C_1 n^{1/2-c_1} \text{ and } \log^{3/2} p \leq C_1 n \tag{27}$$

In addition, suppose that $\inf_{n\geq 1}(\alpha - 2\beta_n) \geq c_1\alpha$ whenever inequality selection is used. Then there exist constants $c, C > 0$ depending only on α, c_1, C_1 such that for every $\epsilon \in (0,1)$, whenever[a]

$$\max_{1\leq j\leq p} \mu_j/\sigma_j \geq (1 + \epsilon + C\log^{-1/2} p)\sqrt{\frac{2\log(p/\alpha)}{n}}$$

we have

$$\mathbb{P}(T > \hat{c}(\alpha)) \geq 1 - \frac{C}{\epsilon^2 \log(p/\alpha)} - Cn^{-c}$$

Therefore, when $p = p_n \to \infty$, for any sequence ϵ_n satisfying $\epsilon_n \to 0$ and $\epsilon_n\sqrt{\log p_n} \to \infty$, as $n \to \infty$, we have

$$\inf_{\mu\in\mathcal{B}_n} \mathbb{P}_\mu(T > \hat{c}(\alpha)) \geq 1 - o(1) \tag{28}$$

where

$$\mathcal{B}_n = \left\{\mu = (\mu_1, \ldots, \mu_p) : \max_{1\leq j\leq p} \mu_j/\sigma_j \geq \bar{r}_n(1 + \epsilon_n)\sqrt{2\log(p_n)/n}\right\}$$

and P_μ denotes the probability measure for the distribution $\mathcal{L}_X$ having mean μ. Moreover, the above asymptotic result (28) holds uniformly with respect to any sequnce of distributions $\mathcal{L}_X$ satisfying (4) and (27).

[a] Seems similar to LASSO condition

Theorem 7 shows that tests are uniformly consistent against all alternatives excluding those in a small neighborhood of alternatives that are too close to the null. Size of this neighborhood is shrikning at a fast rate. They show in a working paper that no test can be consistent against all alternatives whose distance from the null converges to zero faster than $\sqrt{(\log p_n)/n}$ and that the tests above are minimax optimal.

2.6 Monte Carlo Experiments

Results are given, I would check document. This likely concludes my notes on this topic.

[4] Importantly, the three step procedure is not included here.

3 Set Identification in Models with Multiple Equilibria; *Alfred Galichon, Marc Henry (ReStud, 2011)*

Paper proposes a computationally feasible way of deriving the identified features of a model with multiple equilibria in pure or mixed strategies. It can be found from the ReStud website here.

3.1 Introduction

Empirical study of game theoretic models generally complicated by the presence of multiple equilibria. The existence of multiple equiliria generally leads to a failure of identification of the structural parameters governing the model.

- Berry and Tamer (2006) and Ackerberg et al. (2007) give an account of the various ways this issue is approached in the literature.
- Andrews, Berry, and Jia (2003), Ciliberto and Tamer (2009) consider some partial identification approach. Identification approach is not sharp.

Paper proposes a computationally feasible way of recovering the identified. Note a generalized likelihood implied by a model with multiple equilibria can be represented by a non-additive set function called a *Choquet Capacity*. Give a formal definition of an equilibrium selection mechanism and call such a mechanism compatible with the data if the likeli hood of the model augmented with such a mechanism is equal to the probabilities observed in the data. THe identified feature of the model is then the set of parameter values s.t there exists an equilibrium selecton mechanism compatible with the data.

The computational burden remains high in situations with a large number of observable outcomes since the number of inequalities to be checked is equal to the number of subsets of the set of observable outcomes.

3.2 Identified Features of Models with Multiple Equilibria

First go over framework and general results in the case where only equilibria in pure strategies are considered. Section 1.2 specializes and illustrates them on leading examples of participation gaems.

3.2.1 Identified Parameter Sets in General Models with Multiple Equilibria

General framework is that of Jovanovic (1989). Applies to the empirical analysis of normal form games, where only equilibra in pure strategies are considered. Consider three types of economic variables.

- Outcome variables, Y
- Exogeneous eexplanatory variables X
- Random shocks (or latent variables) ϵ

Outcome variables and latent variables are assumed to belong to complete and seperable metric spaces. Economic model consists of a set of restrictions on the joing behavior of the variables listed above. Restrivtions may be induced by assumptions of rational agents, and they generally depend on a set of unknown structural paramaters, θ.

Without loss of generality, model may be formalized as a measurable correspondence (defined below) between the latent variables ϵ and the outcome variables Y, indexed by the exogeneous variables X and the vector of parameters θ. This correspondence is called G and write $Y \in G(\epsilon|X;\theta)$ to indicate admissable values of Y given ϵ, X, θ. The econometrician is assumed to have access to a sample of i.i.d vectors (Y, X) and the problem that is considered is estimating the vector of parameters θ. The latent variables ϵ is supposed to be distributed accorind to a parametric distribution $\nu(\cdot|X;\theta)$. Assumptions are collected below:

Assumption 1. An independent and identically distributed sample of copies of the random vector (Y, X) is available. The observable outcomes Y conditionally distributed accotding to the probability distribution $P(\cdot|X)$ on $\mathcal{Y}$, a Polish space[a] endowed with its Borel σ-algebra of subsets $\mathcal{B}$ are related to the unobervable variables ϵ according to the model $Y \in G(\epsilon|X;\theta)$. Here, θ belongs to an open subset Θ of $\mathbb{R}^{d_\theta}$, ϵ is distributed according to the probability measure $\nu(\cdot|X;\theta)$ on $\mathcal{U}$[b], and G is a measurable correspondence[c] for almost all X and for all $\theta \in \Theta$. Finally, the variables (Y, X, ϵ) are defined on the same underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

[a]A Polish space is a complete and seperable metric space. A complete metric space is one where every Cauchy sequence converges to a point in the space and a seperable metric space is one with a countable, dense, subset. A Cauchy sequence, $(a_n)_{n\in\mathbb{N}}$ is one such that for every $\epsilon > 0, \exists N_\epsilon \in \mathbb{N}$ s.t $\forall m, n \geq N_\epsilon, d(a_m, a_n) < \epsilon$. Completeness ensures that the space is "rich enough." For example, the reals are complete but the rationals are not.

[b]Also a Polish space

[c]A measurable correspondence is such that for all open subsets $A \subseteq \mathcal{Y}$, $G^{-1}(A|X;\theta) := \{\epsilon \in \mathcal{U} : G(\epsilon|X;\theta) \cap A \neq 0\}$ is measurable. A note: technically A should be measurable, but since we are dealing with Borel σ-algebra's this is an equivalent definition. A measurable correspondence is also called a random correspondence or a random set.

Example 1. To illustrate, consider a simple game proposed by Jovanovic (1989). Consider two firms with profit functions $\Pi_1(Y_1, Y_2, \epsilon_1, \epsilon_2; \theta) = (\theta Y_2 - \epsilon_2)Y_1$ and $\Pi_2(Y_1, Y_2, \epsilon, \epsilon_2; \theta) = (\theta Y_1 - \epsilon_1)Y_2$ where $Y_i \in \{0, 1\}$ is firm i's action and $\epsilon = (\epsilon_1, \epsilon_2)$ are exogeneous cots. The firms know thier costs, the analyst only knows that ϵ is uniformly distributed on $[0, 1]^2$ and that the structural parameter θ is in $(0, 1]$. There are two pure strategy Nash equilibria. The first is $Y_1 = Y_2 = 0$ for all $\epsilon \in [0, 1]^2$. The second is $Y_1 = Y_2 = 1$ for all $\epsilon \in [0, \theta]^2$ and $Y_1 = Y_2 = 0$ otherwise. Hence the model is described be the correspondence $G(\epsilon; \theta) = \{(0, 0), (1, 1)\}$ for all $\epsilon \in [0, \theta]^2$ and $G(\epsilon; \theta) = \{0, 0\}$ otherwise.

To conduct inference on the parameter vector θ, one first needs to determine the identified features of the model. Since G may be multi-valued due to the presence of multiple equilibria, the outcomes may not be uniquely determined by the latent variable. In such cases, the generalized likelihood of an outcome falling in the subset A of $\mathcal{Y}$ predicted by the model is $\mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$. Because of multiple equilibria, this generalized likelihood may sum to more than one, as we may have $A \cap B = set$ and yet $G^{-1}(A|X;\theta) \cap G^{-1}(B|X;\theta) \neq 0$[1] so that $\mathcal{L}(A \cup B|X;\theta) < L(A|X;\theta) + L(B|X;\theta)$. The set function $A \mapsto \mathcal{L}(A|X;\theta) = \nu(G^{-1}(A|X;\theta)|X;\theta)$ is generally not additive and is called a *Choquet capacity*[2]. This non-additivity of the model likelihood is well documented.

Definition 1 (Choquet capacity)**.** A Choquet capacity $\mathcal{L}$ on a finite set $\mathcal{Y}$ is a set function $\mathcal{L} : A \subset \mathcal{Y} \mapsto [0, 1]$ which is

- normalized, i.e $\mathcal{L}(\emptyset) = 0$ and $\mathcal{L}(\mathcal{Y}) = 1$
- monotone, i,e $\mathcal{L}(A) \leq \mathcal{L}(B)$, for any $A \subset B \subset \mathcal{Y}$

This is like a probability measure but without additivity. In the example above, $\nu(\cdot|X;\theta)$ is the uniform distribution on $[0, 1]^2$ and the Choquet capacity $\nu(G^{-1})$ gives value $\nu(G^{-1}(\{0, 0\})) = \nu([0, 1]^2) = 1$. and $\nu(G^{-1}\{1, 1\}) = \nu([0, \theta]^2) = \theta^2$ to the set $\{(1, 1)\}$. Hence it is immediately apparent that the Choquet capacity of $\nu(G^{-1})$ is nonadditive.

As discussed in Jovanovic (1989) and Berry and Tamer (2006), the model with multiple equilibria can be completed with an equilibrium selection mechanism. Define an equilibrium selection mechanism as a conditional distribution $\pi_{Y|\epsilon,X;\theta}$ over equilibrium outcomes Y in the regions of multiplicity. By constriction, an equlilibrium slection mechanism is allowed to depend on the latent variables ϵ even after conditioning on X.

Definition 2 (Equilibrium selection mechanism)**.** An equilibrium selection mechanism is a conditional probability $\pi(\cdot|\epsilon, X;\theta)$ for Y conditional on ϵ and X such that the selected value of the outcome variable is actually an equilibrium. Formally $\pi(\cdot|\epsilon, X, \theta)$ has support contained in $G(\epsilon|X;\theta)$.

[1]Inutition: $G^{-1}(y|X;\theta)$ gives the set of ϵ values that could have generated the y value (observed outcome) conditional on X and θ and one epsilon can generate two different y values because of multiple equilibria.

[2]See Choquet, 1954. Choquet capacity also used as a generalized probability in some behavioral decision making theory.

Crucial to this is the fact that π is a *probability measure*. It should "smooth out" the nonadditivity of $\nu(G^{-1})$.

The identified feature of the model is the smallest set of parameters that cannot be rejected by the data. Hence, it is the set of parameters for which one can find an equilibrium selection mechanism that completes the model and equates probabilities of outcomes predicted by the model with the probabilities obtained from the data.[3]

Definition 3 (Compatabile equilibrium selection mechanism)**.** The equilibrium selection mechanism $\pi(\cdot|\epsilon, X;\theta)$ is compatible with the data if the probabilities ovserved in the data are equal to the probabilities predicted by the equilibrium selection mechanism. More formally, if for all measurable subsets A of $\mathcal{Y}$

$$P(A|X) = \int_{\mathcal{U}} \pi(A|\epsilon, X, \theta)\nu(d\epsilon|X;\theta)$$

Definition 4 (Identified Set)**.** The identified set (or the *sharp* identified set) is the set $\theta_I \subseteq \Theta$ such that, $\forall \theta \in \theta_I$, there exists an equilibrium selection mechanism compatible with the data.

Above definition is not operational, in the sense that it does not allow for the computation of the identified set based on the knowledge of the probabilities in the data, because π is an infninite dimensional nuisance parameter. Now set out to show how to reduce the dimensionality of the problem. Equivalent formulation of the identified set relates to the *core* of the Choquet capacity.

Definition 5 (Core of a Choquet capacity)**.** The *core* of a Choquet capacity $\mathcal{L}$ on $\mathcal{Y}$ is the collection of probability distributions Q on $\mathcal{Y}$ such that for all $A \subset \mathcal{Y}$, $Q(A) \leq \mathcal{L}(A)$. [a]

[a]Equivalently, if we consider the random set $\mathcal{L}$ as a map from $(\Omega, \mathcal{F}, \mathbb{P}) \to 2^{\mathcal{Y}}$, we can say a random variable $\gamma : \Omega \to \mathcal{Y}$ is in the core of $\mathcal{L}$ if $\gamma(\omega) \in \mathcal{L}(\omega), \forall \omega \in \Omega$. The random variable γ induces a distribution on $\mathcal{Y}$ that has the above property, and every distribution on $\mathcal{Y}$ with the above property should be induced by a random variable with this property.

In cooperative game theory, a Choquet capacity on a set $\mathcal{Y}$ is interpreted as a game, where $\mathcal{Y}$ is the set of players and $\mathcal{L}$ is the utility value or worth of a coalition $A \subseteq \mathcal{Y}$ and the core of the game $\mathcal{L}$ is the collection of allocations that cannot be improved upon by any coalition of players.

In Example 1, the core of the Choquet capcaity νG^{-1} is the set of probabilities P for the observed outcomes $(0,0)$ and $(1,1)$ such that $P(\{(0,0)\}) \leq \nu G^{-1}(\{(0,0\}) = \nu([0,1]^2) = 1$ and $P(\{(1,1)\}) \leq \nu G^{-1}(\{(1,1)\}) = \nu([0,\theta]^2) = \theta^2$.

Next result shows the equivalende between the existence of a compatible eqm. selection mechanism and the fact that the true distribution of the data belongs to the core of the Choquet capacity that characterizes the generalized likelihood predicted by the model.

Theorem 1. *The identified set Θ_I is the set of parameters such that the true distribution of the observable variables lies in the core of the generalized likelihood predicted by the model.*

$$\Theta_I = \left\{\theta \in \Theta : \forall A \in \mathcal{B}, P(A|X) \leq \mathcal{L}(A|X;\theta); X - a.s\right\}$$

A later theorem generalizes this to the case of mixed strategy eqm. In the example above, the identified set is the set of values for θ such that $0 \leq \mathbb{P}((Y_1, Y_2)) \leq \theta^2$.

The first thing to note from this theorem is that the problem of computing the identified set has been transformed into a finite-dimensional problm in the case where $\mathcal{Y}$ is a finite set. Indeed, in this case, teh problem of computing the indentified set is reduced to the problem of checking a finite number of inequalities.

However, in cases where the cardinality of $\mathcal{Y}$ is large, then the number of inequalities to be checked is $2^{|\mathcal{Y}|} - 2$ and the computational burden is only partially lifted. The rest of the paper is based on the characterization of Theorem 1.

[3]Is this somehow restrictive? I guess not, since we've placed no assumption on the selection mechanism.

3.2.2 Some illustrative examples

Examples are given in this section of Market Entry, Family Bargaining etc. For the most part, they resemble the market entry example given above in Section 2.

Family Bargaining Go over this game since it is later considered in the optimal transport section. Consider a simplified version of the bargaining model of decision regarding the long-term care of an elderly parent for a family with two children. The issue is which child will care for the parent when the parent ages or whether the parent is moved to a nursing home. The payoff to family memeber $i, i = 1, 2$ is represented by the sum of thgree terms.

The first term, V_{ij} represents the value to child i of care option j, wjere $j > 0$ means child j becomes the primary care giver, and $j = 0$ means the parent is moved to a nurding home. The matrix $(V_{ij})_{ij}$ is known to both children. We suppose it takes the form

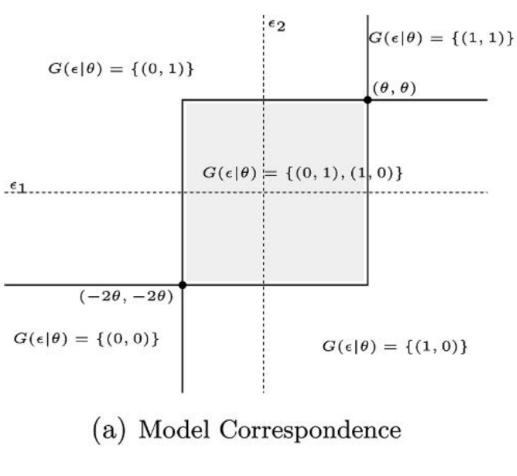
$$V = \begin{pmatrix} 0 & 2\theta & 4\theta \\ 0 & 2\theta & 4\theta \end{pmatrix}$$

$\theta > 0$ is unknown to the analyst. Both children simultaneously decide whether or not to take part in the long-term care decision. Suppose M is the set of children who participate. The option chosen is option j that maximizes the sum $\sum_{i \in M} V_{ij}$ among the available options. It is assumed that participants abide with the decision and that benefits are then shared equally amond the children participating int he decision through a monetary transfer s_i, which is the second term in the children's payoff. Third term ϵ_i is a random benefit from participation, which is 0 for children who decide not to participate and dsitributed according to $\nu(\cdot|\theta)$ for children who participate[4]. All players observe the realization of ϵ, while the analyst knows only its distribution.

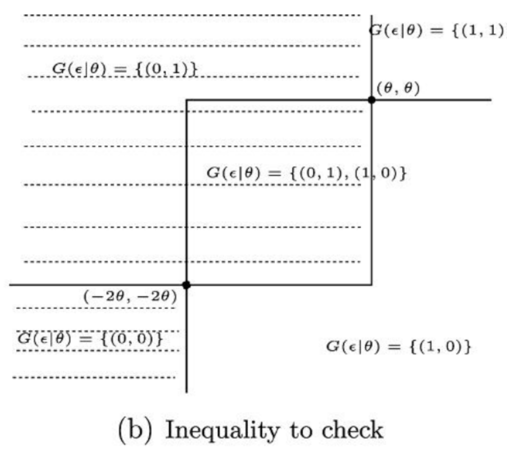
Equilibria correspondence, restricting analysis to only pure strategy NE is:

$$\begin{aligned}
&\{(0,0)\}\text{is a Nash Equilbrium in pure strategies iff } \epsilon_2 < -2\theta \text{ and } \epsilon_1 < -2\theta \\
&\{(1,1)\}\text{is a Nash Equilbrium in pure strategies iff } \epsilon_2 > \theta \text{ and } \epsilon_1 > \theta \\
&\{(0,1)\}\text{is a Nash Equilbrium in pure strategies iff } \epsilon_2 > -2\theta \text{ and } \epsilon_1 < \theta \\
&\{(1,0)\}\text{is a Nash Equilbrium in pure strategies iff } \epsilon_2 < \theta \text{ and } \epsilon_1 > -2\theta
\end{aligned}$$

The equilibrium correspondence $G(\epsilon|\theta)$ is represented in Figure 1(a) below.



(a) Model Correspondence



(b) Inequality to check

Figure 1: Family Game [Lifted from Paper]

[4]Normalizing the utility of the outside option to 0

In the case of the family bargaining game, the set of possible outcomes is $\mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$. The generalized likelihood of outcomes predicted by the model can be written as follows:

$$\begin{aligned}
\mathcal{L}(\{(0,0)\}|\theta) &= \nu(\epsilon : \epsilon_1 \leq -2\theta, \epsilon_2 \leq -2\theta|\theta) = \nu(G^{-1}((0,0)|\theta)|\theta) \\
\mathcal{L}(\{(0,1)\}|\theta) &= \nu(\epsilon : \epsilon_1 \leq \theta, \epsilon_2 \geq -2\theta|\theta) = \nu(G^{-1}((0,1)|\theta)|\theta) \\
\mathcal{L}(\{(1,0)\}|\theta) &= \nu(\epsilon : \epsilon_1 \geq -2\theta, \epsilon_2 \leq \theta|\theta) = \nu(G^{-1}((1,0)|\theta)|\theta) \\
\mathcal{L}(\{(1,1)\}|\theta) &= \nu(\epsilon : \epsilon_1 \geq \theta, \epsilon_2 \geq \theta|\theta) = \nu(G^{-1}((1,1)|\theta)|\theta)
\end{aligned}$$

anf the generalized likelihood of the remaining events can be derived as follows

$$\begin{aligned}
\mathcal{L}(\{(0,0)\} \cup A|\theta) &= \mathcal{L}(\{(0,0)\}|\theta) + \mathcal{L}(A|\theta), \quad \text{for all } A \subset \mathcal{Y}/\{(0,0)\} \\
\mathcal{L}(\{(1,1)\} \cup A|\theta) &= \mathcal{L}(\{(1,1)\}|\theta) + \mathcal{L}(A|\theta), \quad \text{for all } A \subset \mathcal{Y}/\{(1,1)\} \\
\mathcal{L}(\{(0,1),(1,0)\}|\theta) &= 1 - \mathcal{L}(\{(0,0),(1,1)\}|\theta)
\end{aligned}$$

The generalized likelihood predicted by the model is the set function $A \mapsto \mathcal{L}(A|\theta) = \nu(G^{-1}(A|\theta)|\theta)$ for $A \subset \mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$. This set function is a Choquet capacity and if the support of ν is suffeciently large, the generalized likelihood sums to more than one because the region of multiple equilibria is "counted twice".

Model is completed by adding an equilibrium selection mechanism that will pick out a single equilibrium for each value of the latent variable ϵ in the region of multiplicity. As formally defined previously, an equilibrium selection mechanism is a conditional probability $\pi(\cdot|\epsilon, X, \theta)$ with support included in $G(\epsilon|X;\theta)$. It is compatible with the data if the probabilities it predicts are equal to the true probablities of the observable variables.

In this example, for $j = 0, 1$:

$$P((i,j)|X) = \int_{\mathcal{U}} \pi((i,j)|\epsilon, X; \theta)\nu(d\epsilon|X;\theta)$$

Since the model contains no prior information, any valid probability measure equilibrium selection mechanism that generates equates predicted probabilities with observed probabilities is consistent.

It is noted that the definition of the identified region using a semi-parametric likelihood representation, with the equilibrium selection mechanism as the infinite dimensional nuisance parameter π is impractivle, so Theorem 1 is used to make it operational and compute Θ_I. So

$$\Theta_I = \{\theta \in \Theta : (\forall A \in 2^{\mathcal{Y}}; P(A|X) \leq \mathcal{L}(A|X;\theta); X - a.s\}$$

3.3 Effecient Computation of the Identified Set

Subtitle: "*Which inequalities to check and how to check them?*"

Describe three approaches to the effective computation of the identified set based on the characterization of Theorem 1. First approach is based on submodular optimization and extends readily to the case with mixed strategies. Second appraoch, describes in Section 2.2, relies on the highly effecient algorithms for optimal transportation problems.[5] Third approach is based off the notion of *core determining sets* and provides a dramatic reduction in the computational complexity under specific assumptions on the game under study.

3.3.1 Submodular Optimization

The first proposal to deal with the complexity of the problem of checking inequalities in Theorem 1 is a method of general validity based on the minimization of a submodular function, the discrete equivalent of a

[5] I found the following ArXiv introduction to optimal transport problems here (if the link doesn't work; https://arxiv.org/abs/1009.3856)

convex function. This is a well-known problem in combinatorial optimiazation and effecient algorithms are easily available off the shelf.

Definition 6 (Submodular function)**.** A set function $\mathcal{L} : \mathcal{Y} \to \mathbb{R}$ is called submodular if, for each $A, B \subset \mathcal{Y}$, we have

$$\mathcal{L}(A \cup B) + \mathcal{L}(A \cap B) \leq \mathcal{L}(A) + \mathcal{L}(B)$$

In the case that $\mathcal{L}$ is a probability measure, this holds as equality.

Submodularity for set functions is the analogue of convexity, and the problem of minimizing a submodular function is well studied. Paper now shows that checking inequalities involved in the characterization of the identified set in Theorem 1 is equivalent to the minimization of a submodular function. Theorem 1 whoes that the identified set is the set of values of θ such that X-almost surely, we have the domination $\forall A \subseteq \mathcal{Y}$, $P(A|X) \leq \mathcal{L}(A|X;\theta)$. Equivalently,

$$\min_{A \subseteq \mathcal{Y}} \big(\mathcal{L}(A|X;\theta) - P(A|X)\big) \geq 0$$

First note that the function above is indeed submodular.

Lemma 1 (Submodularity of the generalized likelihood)**.** *For all $\theta \in \Theta$ and all X, the set function $\mathcal{Y}$ defined for all $A \subseteq \mathcal{Y}$ by $A \mapsto \mathcal{L}(A|X;\theta) - P(A|X)$ is submodular.*

The most effecient, generic, way to check that a convex function is everywhere non-negative and verify that the minimum is non-negative. Apply the same logic to the above. Of course, can speed this up by terminating the algorithm when a negative value is found.

Theorem 2 (Computation of the identified set)**.** *The identified set is obtained by minimization of a submodular function*

$$\theta_I = \left\{ \theta \in \Theta : \min_{B \subseteq \mathcal{Y}} \big(\mathcal{L}(B|X;\theta - P(B|X)\big) = 0, X - a.s \right\}$$

As a note: I think the reason there is an "$= 0$" instead of a ≥ 0 in the statement of the identified set above is that we can always take $B = \emptyset$. More details on the procedure are given later on in Section 4. This method can be generalized to the case where equilibria in mixed strategies are considered.

The below is a special case of submodular optimization which is more effecient and applies to the case where only equilibria in pure strategies are considered.

3.3.2 Optimal Transportation Approach

When equilibria are only in pure strategies, the model generalized likelihood $\mathcal{L}$ is a very special case of submodular function since it is derived as the distribution function of a random set.

$$\mathcal{L}(A|X;\theta) = \nu(\epsilon : G(\epsilon|X;\theta) \cap A \neq \emptyset|X;\theta)$$

When mixed equilibria are considered, this improvment in effeciency is no longer available because (in general), the model generalized likelihood is no longer the distribution of a random set.[6] To describe the method, need the following notations and definitions.

Call $\mathcal{U}^*$ the set of predicted combinations of equilibrium, formally $\mathcal{U}^* = \{G(\epsilon|X;\theta); \epsilon \in \mathcal{U}\}$, remembering that $\mathcal{U}$ is the support of ϵ. Note that $\mathcal{U}^*$ is a quotient space for the correspondence G [7]. So $\mathcal{U}^*$ contains

[6] Why not? Maybe because the probabilities of observing an outcome now are not functions of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but, rather also depend on the properties of the mixed strategy equilibrium

[7] From WolframMathWorld: A quotient space, $X/\!\sim$ of a topological space X and a set of equivalence classes $\sim$ on X is the set of equivalence classes of points in X (under $\sim$). Open sets on $X/\!\sim$ can be described using the map $\pi : X \to X/\!\sim$ which maps each point in X to its equivalence class. A subset $W \subseteq X/\!\sim$ is open if $\pi^{-1}(W)$ is open. $\mathcal{U}^*$ is a quotient space for $\mathcal{U}$ using equivalence classes from the correspondence G, that is $\mathcal{U} = X/\overset{G}{\sim}$ where $\epsilon_1 \overset{G}{\sim} \epsilon_2$ if $G(\epsilon_1) = G(\epsilon_2)$

subsets of $\mathcal{Y}$ but is typically of much lower cardinality than $2^{\mathcal{Y}}$.

Further, consider the bipartite graph $\mathcal{G}(\theta, X)$ in $\mathcal{Y} \times \mathcal{U}^*$. The edges are defined as $(y, u) \in E(\mathcal{G})$ if $y \in u$. Each vertex $u \in \mathcal{Y}$ has weight $P(y|X)$ and each vertex $y \in \mathcal{U}^*$ has weight $\nu(\{\epsilon : G(\epsilon|X;\theta) = u|X\})$. Finally, call $Q(\cdot|X;\theta)$ the probabilities $Q(u|X;\theta) = \nu\{G^{-1}(u)|X;\theta)\}$ (so t at $Q(u|X;\theta)$ is the weight attached to vertex $u \in \mathcal{Y}$.)

Theorem 1 shows that $\theta \in \Theta_I$ if and only if, for any subset A of $\mathcal{Y}$, we have $P(A|X) \leq Q(G^{-1}(A)|X;\theta)$, where $G^{-1}(A) = \{u \in \mathcal{U}^* | A \cap u \neq \theta\}$. Galichon and Genry show that it is equivalent to the existance of a joint probability Λ on $\mathcal{G}(\theta, X)$ with marginal distributions $P(\cdot|X)$ and $Q(\cdot|X;\theta)$.[8]

Theorem 3. *The parameter value θ belongs to the identified set iff there exists a probability on $\mathcal{Y} \times \mathcal{U}^*$ with support contained in $G(X;\theta)$ and with marginal probabilities $P(\cdot|X)$ and $Q(\cdot|X;\theta)$.*

One implication is easy to prove. Call U the random element with distribution Q (U is a random variable from $(\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{Y}$) If a joint probability Λ exists with all the required properties then

$$Y \in A \implies U \in G^{-1}(A)$$

(Family Bargaining Example cont.) For the case of the family bargaining game

$$\mathcal{U}^* = \Big\{\{(0,0)\}, \{(0,1)\}, \{(1,0)\}, \{(1,1)\}, \{(0,1), (1,0)\}\Big\}$$

This is a class of sets because $G(\epsilon|X;\theta)$ is a correspondence. The existence of a joint probability measure on $\mathcal{Y} \times \mathcal{U}^*$ supported on $\mathcal{G}(X;\theta)$ with marginal probabilities $p_y, y \in \mathcal{Y}$ and $q_u, u \in \mathcal{U}^*$ can be represented graphically by a set of non-negative numbers attached to each edge of the graph that sum to 1, and such that the weight of each vertex is qual to the sum of the weights on the edges that reach it.

[8] I'm a bit lost on this part. I don't quite know what a probability is on a bipartite graph. I guess we just mean a joint probability distribution on the product space of the outcomes and the sets of outcomes. The theorem makes more sense intuitively looking at it from this view. For a consistent parameter, the rules relating the weights on one side with weights on the other are given above.