June 30, 2020

Tuning parameter choice

This document describes the way in which an important tuning parameter was chosen. Data was generated from two normal distributions, labeled with a 0 or a 1. Distribution 0 had mean zero and distribution 1 had mean μ that was allowed to shift, and both had scale equal to one. For a pair of standard normal distributions the true lower bound function has a convenient functional form:

$$L_0(\delta) = \begin{cases} 0 & \delta \leq 0 \\ 2\Phi(\delta/2) - 1 & \delta > 0 \end{cases}, \tag{1}$$

where Φ is the standard normal CDF. This was the null hypothesis, while data was generated under both the null and locally location-shifted alternative distributions. Specifically, the location parameter of distribution 1, μ, was allowed to range from $-5/\sqrt{n}$ to $5/\sqrt{n}$, with $n = n_0 + n_1$, where n_0 and n_1 are the sample sizes from each distribution. In simulations both samples had the same size, and the sample sizes used were $n = 100, 500$ and 1000. Samples were generated from these two distributions and the empirical lower bound was calculated:

$$\mathbb{L}_n(\delta) = \max_y \mathbb{F}_{1n}(y) - \mathbb{F}_{0n}(y - \delta).$$

Then we subtract the null L_0 and scale the difference to produce an empirical bound process.

Because it is difficult to evaluate the empirical difference process at each point where it changes, the empirical difference process $\sqrt{n}(\mathbb{L}_n - L_0)$ was evaluated on a grid $\mathbb{X}$ of points ranging from $\min_i X_{i1} - \max_j X_{j0}$ to $\max_i X_{i1} - \min_j X_{j0}$ with a mesh diameter 0.05.

For bootstrapping, we estimate the Hadamard directional derivative of the pointwise supremum map, and this requires estimates of the marginal ϵ-maximizer sets that depend on a tuning parameter labeled a_n in the text. We chose $a_n = K \log(\log(n))/n$ and chose K using the results of these experiments. To be explicit, we estimate a choice set for the derivative using, for each $x \in \mathbb{X}$,

$$U_{f_n}(x, a_n) = \left\{ u \in \mathbb{R} : \mathbb{F}_{1n}(u) - \mathbb{F}_{0n}(u - x) \geq \max_u (\mathbb{F}_{1n}(u) - \mathbb{F}_{0n}(u - x)) - a_n \right\}$$

Then bootstrap samples are used to construct the statistic, with $h_r^*(u, x) = \mathbb{F}_{1n}^*(u) - \mathbb{F}_{0n}^*(u - x) - (\mathbb{F}_{1n}(u) - \mathbb{F}_{0n}(u - x))$, for $r = 1, \ldots, R$,

$$\lambda_{1r}^* = \hat{\lambda}_{1n}'(h_r^*) = \max \left\{ \sup_{(u,x) \in U_{f_n}(X, a_n)} h_r^*(u, x), \sup_{x \in X} \inf_{u \in U_{f_n}(x, a_n)} (-h_r^*(u, x)) \right\}$$

or

$$\lambda^*_{3r} = \hat{\lambda}'_{3n}(h^*_r) = \left(\int_X \left| \sup_{u \in U_{f_n}(x, a_n)} h^*_r(u, x) \right|^p \mathrm{d}m(x) \right)^{1/p}.$$

The results below show that the supremum norm statistics were very close to the desired 5% rejection probability, while the empirical rejection probabilities of L^2 statistics were a little below the target.

In the below tables, the row labeled "0" corresponds to when data is really generated from two standard normal distributions, and is the case when L_0 is the correct centering for the sample analog. All the non-zero rows are local alternatives $\mu/\sqrt{n}$, although they are labeled only for the value μ that was used. From the results presented here, we decided to use $K = 0.2$ in the main paper.

Table 1: Rejection probabilities: sample size 100

μ	KS				CvM			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
-5	0.938	0.949	0.931	0.919	0.957	0.946	0.918	0.901
-4	0.806	0.812	0.798	0.810	0.880	0.831	0.784	0.753
-3	0.597	0.634	0.633	0.615	0.704	0.670	0.596	0.562
-2	0.389	0.388	0.372	0.356	0.492	0.423	0.333	0.300
-1	0.175	0.174	0.155	0.161	0.253	0.201	0.156	0.143
0	**0.061**	**0.083**	**0.052**	**0.064**	**0.087**	**0.088**	**0.051**	**0.043**
1	0.014	0.018	0.023	0.026	0.029	0.028	0.015	0.016
2	0.031	0.028	0.020	0.016	0.061	0.031	0.017	0.008
3	0.089	0.071	0.084	0.075	0.169	0.096	0.090	0.058
4	0.202	0.244	0.220	0.206	0.303	0.299	0.235	0.171
5	0.436	0.474	0.428	0.442	0.578	0.529	0.443	0.404

Table 2: Rejection probabilities: sample size 500

μ	KS				CvM			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
-5	0.925	0.922	0.921	0.915	0.950	0.928	0.907	0.886
-4	0.771	0.811	0.760	0.783	0.806	0.802	0.725	0.716
-3	0.588	0.581	0.588	0.538	0.648	0.603	0.559	0.463
-2	0.347	0.381	0.350	0.296	0.400	0.373	0.304	0.226

-1	0.171	0.158	0.144	0.136	0.198	0.159	0.113	0.099
0	**0.063**	**0.053**	**0.054**	**0.044**	**0.074**	**0.052**	**0.042**	**0.033**
1	0.024	0.019	0.025	0.028	0.034	0.031	0.015	0.016
2	0.042	0.054	0.041	0.051	0.063	0.060	0.033	0.042
3	0.122	0.131	0.120	0.106	0.197	0.148	0.117	0.107
4	0.307	0.300	0.301	0.270	0.413	0.355	0.322	0.248
5	0.521	0.533	0.538	0.511	0.654	0.604	0.554	0.497

Table 3: Rejection probabilities: sample size 1000

μ	KS				CvM			
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
-5	0.913	0.925	0.923	0.906	0.930	0.918	0.903	0.873
-4	0.774	0.806	0.789	0.752	0.817	0.793	0.755	0.681
-3	0.575	0.558	0.553	0.529	0.622	0.549	0.527	0.448
-2	0.351	0.324	0.313	0.309	0.375	0.318	0.260	0.252
-1	0.154	0.167	0.153	0.142	0.182	0.155	0.123	0.102
0	**0.057**	**0.050**	**0.043**	**0.049**	**0.074**	**0.046**	**0.033**	**0.033**
1	0.026	0.018	0.020	0.017	0.040	0.027	0.019	0.013
2	0.051	0.053	0.055	0.052	0.084	0.078	0.061	0.047
3	0.155	0.156	0.132	0.121	0.236	0.190	0.139	0.122
4	0.330	0.352	0.332	0.304	0.430	0.403	0.369	0.290
5	0.580	0.574	0.586	0.549	0.711	0.653	0.599	0.551

Setting $K = 0.2$ appeared to make most of the empirical rejection probabilities close to the target under the null hypothesis for the larger sample sizes.