ALTERNATIVE EMPIRICAL DISTRIBUTIONS BASED ON WEIGHTED LINEAR COMBINATIONS OF ORDER STATISTICS

Theodore P. Hill1 and James Mann
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT
A class of empirical distributions is introduced which are based on various weighted linear combinations of order statistics, and which have convergence properties the classical empirical distribution does not, or which stochastically or convexly dominate the classical empirical distribution.

§1. INTRODUCTION
In estimating an unknown probability distribution $F$ based on a sequence of iid observations $X_1, X_2, \ldots$ (with order statistics $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$) from $F$, it is usually desirable to place positive mass only on the observations seen, as does the classical empirical distribution (random probability measure)

$$F_n = \sum_{i=1}^{n} n^{-1} \delta(x_i) = \sum_{i=1}^{n} n^{-1} \delta(x_{i:n})$$

(where $\delta(x)$ is the one-point Dirac measure of mass 1 at $\{x\}$); to place mass on unobserved values seems artificial. However, the relative frequency of values observed in any given interval varies stochastically depending on the observations $X_1, \ldots, X_n$, so weighting the observations exactly uniformly is

---

1 Research partially supported by U.S. NSF Grants DMS 92-03524 and 95-03119 and by Dutch NSF (NWO) Grant Dossier B-61-355.
not really crucial. For example, in sampling from a Bernoulli distribution with \( p = 1/2 \) (fair-coin tossing), after \( n \) tosses only values of 0 and 1 will have been observed, but the exact relative frequencies oscillate randomly about 1/2 (with oscillations described by the law of the iterated logarithm). Thus if the exactly uniform weights \( 1/n \) of the classical empirical distribution are replaced by weights close to \( 1/n \), the result will be qualitatively the same, namely, only the observed values of 0 and 1 will have positive mass, and these masses will be close to 1/2 each.

If these deviations from the uniform masses \( 1/n \) have appropriate structures, various properties not shared by the classical empirical distribution may be attained, such as convergence of the quantile functions, one-sided convergence of means, stochastic or convex domination of \( F_n \), correct means or medians when those are known, or various smoothness properties, in addition to convergence to the underlying \( F \) uniformly almost surely.

The main goal of this article is to introduce a large class of such alternative empirical distribution functions, thus extending the idea underlying the one-sided empirical distributions in [1], [3], [4]. As in those empirical distributions, the ones introduced here are based on nonuniform weights on the order statistics of the random sample, and as such complement empirical distributions based on nonuniform weights on the unordered sample elements (e.g. [6], [9]), on U-statistics structures (cf. [5]), on maximum likelihood properties, (cf. [7]), and ones based on given sufficient statistics (Rao-Blackwellization method), the method of kernels, and the method of stochastic approximation (cf. [8], Chapter 9). After the general class is described, a number of examples and specific results are proved.

§2. EMPIRICAL WEIGHT FUNCTIONS AND DISTRIBUTION SELECTORS

The alternative empirical distributions described in this section are natural generalizations of the classical empirical distribution \( F_n \), having the form

\[
D_n = \sum_{i=1}^{n} w_{i:n} \delta(X_{i:n})
\]

where the weights \( \{w_{i:n}\} \) are nonnegative random variables which depend only on the order statistics \( \{X_{i:n}\} \), and which sum to 1. In order to guarantee
that these weights do not depend on specific distributions of \( \{X_{i:n}\} \), a general measurable selection setting will be used for each fixed \( n \), and these will then be strung together to form a general selection procedure for generating empirical distributions for all \( n \).

Note: The same symbol will be used to denote both a Borel probability measure on \( \mathbb{R} \) and its distribution function; e.g., depending on context, \( F \) is a probability measure or is the corresponding cumulative distribution function, so \( F(x) = F(-\infty, x] \), and \( F(\{x\}) \) is the measure of the (singleton) point \( x \).

**Definition 2.1.** An empirical weight selector for a sample of size \( n \) is a permutation-invariant function \( w_n = (w_{1:n}, \ldots, w_{n:n}) : \mathbb{R}^n \rightarrow [0, 1]^n \) satisfying \( \sum_{i=1}^n w_{i:n} = 1 \).

Thus each \( w_{i:n} \) is a Borel function from \( \mathbb{R}^n \) to \( [0, 1] \) satisfying

\[
w_{i:n}(r_1, \ldots, r_n) = w_{i:n}(r_{1:n}, \ldots, r_{n:n}),
\]

where \( r_{1:n} \leq \cdots \leq r_{n:n} \) are the order statistics of \( r_1, \ldots, r_n \); intuitively, \( w_{i:n} \) is simply the weight assigned to the \( i^{th} \) order statistic.

**Example 2.2.**

(i) \( w_{i:n} = \frac{1}{n} \). This corresponds to the uniform weights of the classical empirical distribution \( F_n \).

(ii) \( w_{i:n} = \frac{1}{n} - \frac{n+1}{2n^{\alpha/4}} + \frac{1}{n^{\alpha/4}} \). This is the special case \( \alpha = 9/4 \) of [3], where large order statistics are weighted more heavily than smaller, forming an arithmetic sequence. It is easily checked that these weights are nonnegative and sum to 1, and that the weights are both scale and translation invariant, since \( w_{i:n} \) depends only on \( i \) and \( n \).

(iii) \( w_{i:n} = \frac{n^{-1} - \alpha sgn(r_{i:n})}{1 - d \sum_{i=1}^n sgn(r_{i:n})} \), where \( d = n^{-1} \sum_{i=1}^n r_i / \sum_{i=1}^n |r_i| \) if \( |\sum r_i| < \sum |r_i| \) (and = 0 otherwise). This weight selector assigns mass \((n^{-1} + d)/(1-ds)\) to each observation less than zero, and mass \((n^{-1} - d)/(1-ds)\) to each observation greater than zero, where \( s = \sum |r_i| \) and \( sgn(x) = -1 \) if \( x < 0 \), \( = 0 \) if \( x = 0 \), \( = +1 \) if \( x > 0 \). It is easily seen that these weights are positive-scale invariant, but not translation invariant.

(iv) \( w_{i:n} = (n^{-1} + |r_{i:n}|)/(1 + \sum_{i=1}^n |r_i|) \). These weights, which depend both on the magnitude and position of the observations, are neither scale nor translation-invariant.
Definition 2.3. An empirical distribution selector (e.d.s.) $\mathcal{D}$ is a sequence of functions $\mathcal{D} = \{D_1, D_2, \ldots \}$, where for each $n$, $D_n : \mathbb{R}^\infty \to \{\text{Borel probability measures on } \mathbb{R}\}$ is given by

$$D_n(r_1, r_2, \ldots) = \sum_{i=1}^{n} w_{i:n}(r_1, \ldots, r_n)\delta_{(r_i:n)}$$

for some fixed empirical weight selector $w_n$. (In other words, $D_n$ assigns to each sequence $r_1, r_2, \ldots$ the discrete probability measure on $\mathbb{R}$ having atoms of mass $w_{i:n}$ at $r_{i:n}$ for each $i, 1 \leq i \leq n$.)

Observe that for each sequence of random variables $\mathbf{X} = (X_1, X_2, \ldots)$ on a probability space $(\Omega, \mathcal{F}, P)$, $D(\mathbf{X})$ is a sequence of random probability measures, where for each $n$, $D_n(\mathbf{X})$ is a $\sigma(X_1, \ldots, X_n)$-measurable map from $\Omega$ to finitely-supported (Borel) probability measures on $\mathbb{R}$, and that

$$\text{supp } D_n(\mathbf{X}) \subset \{X_1, \ldots, X_n\} \quad \text{for each } n.$$

Example 2.4.

(i) $\mathcal{D} = (F_1, F_2, \ldots)$ is the classical e.d.s.

(ii) $\mathcal{D}^+ = (D_1^+, D_2^+, \ldots)$, where $w_{i:n}^+$ are the weights in Example 2.2(ii). It is easily seen [3] that

$$D_n^+ = F_n - 2n^{-1/4}F_n(1 - F_n).$$

(iii) $\mathcal{D}^- = (D_1^-, D_2^-, \ldots)$ where $w_{i:n}^-$ are the corresponding decreasing weights $w_{i:n}^- = \frac{1}{n} + \frac{n^{-1}}{2n^{3/4}} - \frac{i}{n^{3/4}}$, and

$$D_n^- = F_n + 2n^{-1/4}F_n(1 - F_n)$$

(iv) $\mathcal{D} = (D_1, D_2, \ldots)$, where $D_n = F_n + \epsilon_nF_n(1 - F_n)$ for $\epsilon_n \leq 1$ (cf. [1]), or more generally (with $D_n, F_n$ viewed as distribution functions) $D_n = \phi(F_n)$, where $\phi : [0, 1] \to [0, 1]$ is a nondecreasing continuous function which fixes 0 and 1. (All these are generalizations of (ii) and (iii).)

(v) $\mathcal{D} = (D_1, D_2, \ldots)$, where $D_n = F_n$ if $n$ is even, and $D_n = D_n^+$ if $n$ is odd. In this contrived e.d.s. the $\{D_n\}$ are not of the same form for each $n$ and e.d.s.'s of this type will not be studied in this article.
Definition 2.5. An e.d.s. \( D = (D_1, D_2, \ldots) \) is GC (Glivenko-Cantelli) regular if for all \( F \) and all sequences \( X = X_1, X_2, \ldots \) of independent \( F \)-distributed random variables (on \((\Omega, \mathcal{F}, P))\),

\[
\lim_{n \to \infty} ||D_n(X) - F||_{\infty} = 0 \quad \text{P a.s.,} \tag{5}
\]

(\(\|G\|_{\infty} = \sup_x G(-\infty, x]\) is the usual sup-norm on distribution functions).

Theorem 2.6. \( D, D^+ \) and \( D^- \) are GG regular.

Proof. That \( D \) is GC regular is just the Glivenko-Cantelli theorem; the GC regularity of \( D^+ \) and \( D^- \) follows by Theorem 3.3 of [3]. \( \square \)

Example 2.7. \( D = (D_1, D_2, \ldots) \) where \( D_n = 2F_n - F_n^2 \) (this is Example 2.4(iv) with \( \epsilon_n \equiv 1 \)), is not in general GC regular; it converges to the distribution \( 2F - F^2 \).

Although it is this classical strong GC regularity which will be the focus of this article, other notions of regularity are also of interest in this context: CLT (Central Limit Theorem) regularity, which requires that \( \sqrt{n}(D_n(X) - F) \) converge in distribution to a Brownian Bridge (cf. [9]); or regularities which require that the empirical distribution converges to the true distribution with respect to some other metric, such as Kolmogorov-Smirnov or Wasserstein.

§3. KNOWN MEANS, MEDIANS OR MODES

The purpose of this section is to give examples of empirical distribution selectors which incorporate additional information about the underlying distribution \( F \), and which still converge Glivenko-Cantelli-wise to \( F \). In many real-life experiments, parameters about an unknown distribution are sometimes known. For example, the distribution of roundoff errors in scientific calculations is often unbiased (i.e., has mean 0), and in many experiments concerning measurements, the variance is often known as a function of the accuracy of the measuring device. In these cases an empirical distribution is sought which converges to the true distribution, but which in addition has the correct known mean or variance.
For the first example, suppose that $F$ has known finite mean $\mu$. The classical empirical distributions $F_n$ in general almost surely never has mean $\mu$, but certain empirical distributions $D_n$ will now be given which (eventually a.s.) have the correct mean $\mu$, and are still supported by the observations $X_1, \ldots, X_n$ and converge to $F$. There are many such empirical distribution selectors of rather different structures, of which those in the following two theorems are examples.

**Definition 3.1.** For fixed $\mu \in \mathbb{R}$, let $D^{(1)}_{\mu} = (D_1, D_2, \ldots)$ be the e.d.s. with empirical weight functions

$$w_{i:n} = \frac{n^{-1} - d \text{sgn}(r_{i:n} - \mu)}{1 - d \sum_{i=1}^{n} \text{sgn}(r_{i:n} - \mu)}$$

where $d = d(r_1, \ldots, r_n) = n^{-1} \sum_{i=1}^{n} (r_i - \mu) / \sum_{i=1}^{n} |r_i - \mu|$ if $|\sum(r_i - \mu)| < \sum |r_i - \mu|$, and = 0 otherwise. (This is a generalization of Example 2.2(iii) where again the weights assigned to observations less than $\mu$ are all equal, and those assigned to observations bigger than $\mu$ are all equal, but not in general the same as those below.)

Recall that "$E_n$ eventually a.s." means that $P(\lim \inf_{n \to \infty} E_n) = 1$, i.e., for $P$-almost all $\omega \in \Omega$, there exists an $n$ such that $\omega \in E_m$ for all $m \geq n$.

**Theorem 3.2.** For every iid sequence of random variables $\mathbf{X}$ with finite mean $\mu$ and distribution $F$, $D^{(1)}_{\mu} (= (D_1, D_2, \ldots))$ satisfies (5) and

$$D_n(X) \text{ has mean } \mu \text{ eventually a.s.}$$

**Remark.** Note that $D^{(1)}_{\mu}$ is not GC regular since it may not satisfy (5) for $F$ which do not have mean $\mu$, in contrast to the e.d.s $D^{(2)}_{\mu}$ below which is GC regular (for all $F$), but for which (7) requires more than existence of first moment.

**Proof of Theorem.** That the $\{w_{i:n}\}$ defined in (6) form a sequence of empirical weight selectors (nonnegative, sum to 1, Borel measurable, functions only of the order statistics $\{r_{i:n}\}$) is easy to check.

Let $F$ be a distribution with finite mean $\mu$, and let $\mathbf{X} = (X_1, X_2, \ldots)$ be an iid sequence of $F$-distributed random variables. If $F$ is degenerate
(i.e., $X_1 = \mu$ a.s.), the conclusion of the theorem is trivial, since then
\[ d(X_1, \ldots, X_n) = 0 \text{ a.s.,} \quad w_{i:n} = \frac{1}{n} \text{ a.s., and } \mathbb{D}_n = \mathbb{F}_n = \text{the Dirac measure at } \mu \text{ a.s.} \]

Suppose $F$ is not degenerate. By the strong law of large numbers it follows easily that
\[ d(X_1, \ldots, X_n) = o(n^{-1}) \text{ a.s.,} \tag{8} \]
so \[ \|\mathbb{D}_n(\mathbf{X}) - \mathbb{F}_n(\mathbf{X})\|_\infty \leq n \max_{1 \leq i \leq n} \left| n^{-1} - \frac{n^{-1} - d \text{ sign}(X_{i:n} - \mu)}{1 - d \sum \text{ sign}(X_{i:n} - \mu)} \right| = n \left| n^{-1} - \frac{n^{-1} + o(n^{-1})}{1 + o(1)} \right| = o(1) \text{ a.s., and (5) follows by the triangle inequality and Glivenko-Cantelli theorem, } \|\mathbb{F}_n(\mathbf{X}) - F\|_\infty \to 0 \text{ a.s.} \]

To establish (7), first note that since $F$ has mean $\mu$ and is not degenerate, it follows that \[ \sum_{i=1}^n (X_i - \mu) < \sum_{i=1}^n |X_i - \mu| \text{ eventually a.s., so} \]
\[ d(X_1, \ldots, X_n) = n^{-1} \sum_{i=1}^n (X_i - \mu) / \sum_{i=1}^n |X_i - \mu| \text{ eventually a.s.} \tag{9} \]

The mean of $\mathbb{D}_n(\mathbf{X})$ is
\[ \sum_{i=1}^n X_{i:n} w_{i:n}(X_1, \ldots, X_n) = \sum_{i=1}^n X_{i:n} \left[ \frac{n^{-1} - d(X_1, \ldots, X_n) \text{ sign}(X_{i:n} - \mu)}{1 - d(X_1, \ldots, X_n) \sum_{i=1}^n \text{ sign}(X_{i:n} - \mu)} \right] \]
which by (9) equals $\mu$ eventually a.s. \hfill \Box

**Definition 3.3.** Let $\mathbb{D}^+ = (\mathbb{D}_1^+, \mathbb{D}_2^+, \ldots)$ and $\mathbb{D}^- = (\mathbb{D}_1^-, \mathbb{D}_2^-, \ldots)$ be the e.d.s.'s of Example 2.4(ii) and (iii), respectively, and for fixed $\mu \in \mathbb{R}$, let $\mathbb{D}^{(2)}\mu = (\mathbb{D}_1, \mathbb{D}_2, \ldots)$ be the e.d.s. given by
\[ \mathbb{D}_n = \lambda_n \mathbb{D}_n^+ + (1 - \lambda_n) \mathbb{D}_n^- \tag{10} \]
where $\lambda_n : \mathbb{R}^\infty \to [0, 1]$ is defined by
\[ \lambda_n(r_1, r_2, \ldots) = \left( \mu - \sum_{i=1}^n r_{i:n} w_{i:n}^{-}(r_1, \ldots, r_n) \right) / \sum_{i=1}^n r_{i:n} (w_{i:n}^+(r_1, \ldots, r_n) - w_{i:n}^{-}(r_1, \ldots, r_n)) \]
if this quotient is in $[0, 1]$, and $= 1/2$ otherwise. (Intuitively, if the barycenter (mean) of $\mathbb{D}_n^-$ is $\leq \mu$ and that of $\mathbb{D}_n^+$ is $\geq \mu$, then $\lambda_n$ is a convex combination of $\mathbb{D}_n^+$ and $\mathbb{D}_n^-$ which has mean $\mu$. Note that if $\lambda = 1/2$, then $\mathbb{D}_n = \mathbb{F}_n$, the classical empirical distribution.)

**Theorem 3.4.** For every $\mu \in \mathbb{R}$, $\mathbb{D}^{(2)}\mu = (\mathbb{D}_1, \mathbb{D}_2, \ldots)$ is GC regular. In addition, if $\mathbf{X}$ is any iid sequence of random variables having distribution
\[ F, \text{mean} \mu, \text{and finite variance}, \text{then } \mathcal{D}_n(X) \text{ has mean } \mu \text{ eventually almost surely.} \]

**Proof.** By Theorem 3.3 of [3], \( \mathcal{D}^+ \) and \( \mathcal{D}^- \) are both GC regular, so any convex combination such as \( \mathcal{D}^{(2)}_\mu \) also is.

Suppose \( X \) is an iid sequence of random variables having distribution \( F \), mean \( \mu \), and finite second moment (in fact finite \( 4/3 \)-moment will do). By Theorem 2.2 of [3], the mean of \( \mathcal{D}^+_n(X) \) is eventually a.s. \( \geq \mu \), and that of \( \mathcal{D}^-_n(X) \) is eventually a.s. \( \leq \mu \), so eventually a.s. \( \lambda_n(X) \in [0, 1] \), which implies that \( \mathcal{D}^{(2)}_\mu \) has mean \( \mu \) eventually a.s. \( \square \)

These previous two e.d.s.'s are alternatives to the classical empirical distribution when the mean of the underlying distribution is known and analogous constructions yield e.d.s.'s for distributions with known variances; the next is an analog for known median (other quantiles are similar), whose construction is similar to that for \( \mathcal{D}^{(2)}_\mu \).

**Definition 3.5.** Let \( \mathcal{D}^+_n, \mathcal{D}^-_n \) be as above, and for \( h \in \mathbb{R} \), let \( \mathcal{D}^{(3)}_n = (\mathcal{D}_1, \mathcal{D}_2, \ldots) \) be the e.d.s. given by

\[
\mathcal{D}_n = \lambda_n \mathcal{D}^+_n + (1 - \lambda_n) \mathcal{D}^-_n,
\]

where \( \lambda_n : \mathbb{R}^\infty \to [0, 1] \) is defined by

\[
\lambda_n(r_1, r_2, \ldots) = \left( \frac{1}{2} - \sum \{ w^-_{i,n}(r_1, \ldots, r_n) : r_{i:n} \leq h \} \right) / \sum \{ (w^-_{i,n}(r_1, \ldots, r_n) - w^+_{i,n}(r_1, \ldots, r_n)) : r_{i:n} \leq h \}
\]

if this quotient is in \([0, 1]\) and \( = 0 \) otherwise.

(Intuitively, if the total mass assigned to points \( \leq h \) is \( \geq \frac{1}{2} \) for \( \mathcal{D}^-_n \) and is \( \leq \frac{1}{2} \) for \( \mathcal{D}^+_n \), then \( \lambda_n \) is a convex combination of \( \mathcal{D}^-_n \) and \( \mathcal{D}^+_n \) with total mass on points \( \leq h \) exactly equal to \( \frac{1}{2} \) in which case \( h \) is a median of this convex combination.)

**Theorem 3.6.** For any \( h \in \mathbb{R} \), \( \mathcal{D}^{(3)}_n = (\mathcal{D}_1, \mathcal{D}_2, \ldots) \) is GC regular, and if \( X \) is any iid sequence of random variables having distribution \( F \) with median \( h \), then

\[
h \text{ is a median of } \mathcal{D}_n(X) \text{ eventually a.s.}
\]

(12)
Proof. The GC regularity follows as in the proof of Theorem 3.4, and (12) also follows easily by Theorem 3.3 of [3], since for $D_n^+$ (where the larger weights are placed on higher order statistics), the total mass assigned to points $\leq h$ (the true median) is eventually a.s. $\leq 1/2$, and that for $D_n^-$ is eventually a.s. $\geq 1/2$.

There are also e.d.s.'s (similar to $D^{(1)}_n$) which place equal weights on all points larger than $h$, and equal weights on points less than $h$, and which have the correct median eventually a.s.; such constructions are left to the interested reader. The final theorem in this section is an analogous conclusion for known modes (for distributions with at least one atom).

Theorem 3.7. $D^+$ and $D^-$ are both GC regular, and if $F$ is a distribution with at least one atom, then for every iid sequence $X$ with distribution $F$, the mode of $D_n^+(X)$ is eventually a.s. the largest mode of $F$, and that of $D_n^-(X)$ is eventually a.s. the smallest mode of $F$. Moreover, if $F$ is any distribution with finite mean $\mu$, then the means of $D_n^+(X)$ and $D_n^-(X)$ converge to $\mu$ a.s.

Proof. Similar, using Theorem 3.3 (iv) and (v) of [3].

Remarks. The choice of exponent $(n^{3/4})$ for $D_n^+$ and $D_n^-$ was somewhat arbitrary for the median and mode conclusions; any exponent in $(2, 5/2)$ will do. The requirement for $F$ to have at least one atom is to guarantee that the notion of “mode” is unambiguous.

§4. STOCHASTIC DOMINATION, CONVEX DOMINATION, AND SMOOTHNESS

In some situations, when estimating an unknown distribution $F$ based on iid observations $X_1, X_2, \ldots$ from that distribution it might be desirable to use an empirical distribution $D_n$ which is stochastically larger than the classical empirical distribution $F_n$, or which convexly dominates $F_n$, has smaller variance than $F_n$, or is “smoother” than $F_n$ (and which still converges to $F$ uniformly almost surely). There are many such empirical distributions, and in this section several will be described.

As is the main conclusion of Theorem 3.3 of [3], the e.d.s. $D^+$ of Example 2.4(ii) is not only GC regular, and $D_n^+(X)$ is (trivially) stochastically larger than $F_n(X)$ for all $n$ and all iid $X$, but also even the (upper) quantile
functions (inverses) of \( D_n^- (x) \) converge to the (upper) quantile function of \( F \) almost surely and ([3], Theorem 2.2) if \( F \) has finite variance, the means of \( D_n^+ (x) \) converge to the mean of \( F \) \textit{eventually from above} almost surely; \( D^- \) similarly gives a stochastically smaller estimate of \( F \), convergence of lower quantiles and convergence of means from below. In general, stochastic domination of \( F \) itself by empirical distributions is not possible, as can be seen by looking at any \( F \) with unbounded support (cf. also Example 3.4 of [3]).

To attain an e.d.s. \( D = (D_1, D_2, \ldots) \) for which \( D_n (x) \) convexly dominates \( F_n (x) \) for all \( n \) and all iid \( x \) (recall that a (real Borel) probability measure \( G_1 \) dominates a probability measure \( G_2 \) convexly if \( \int f dG_1 \geq \int f dG_2 \) for all convex functions \( f : \mathbb{R} \rightarrow \mathbb{R} \)), it is only necessary to replace \( F_n \)'s uniform weights \( \frac{1}{n} \) at \( r_{i:n} \) with weights \( w_{i:n} \) at \( r_{i:n} \) which form a \textit{balayage} (cf. [2]) of \( F_n (x) \) for all \( n \) and \( x \). The following is a simple typical example in which the weights at the extrema \( X_{1:n} \) and \( X_{n:n} \) are increased, and the weights on interior order statistics decreased uniformly, so that the new measure is a balayage of \( F_n \).

**Definition 4.1.** Let \( D^c = (D_1, D_2, \ldots) \) be the e.d.s. with empirical weight functions

\[
\begin{align*}
w_{1:n} &= n^{-1} + \delta_1, \quad w_{2:n} = \cdots = w_{n-1:n} = n^{-1} - (n \log n)^{-1}, \\
w_{n:n} &= n^{-1} + \delta_2,
\end{align*}
\]

where \( \delta_1 = (r_1, \ldots, r_n) = (n \log n)^{-1} \left( (n-2)r_{n:n} - \sum_{i=2}^{n-1} r_{i:n} \right) / (r_{n:n} - r_{1:n}) \), if \( r_{n:n} - r_{1:n} > 0 \), and \( = 0 \) otherwise, and \( \delta_2 = (n-2)(n \log n)^{-1} - \delta_1 \). (Intuitively, mass of \((n \log n)^{-1}\) each is removed from each of the interior order statistics \( r_{2:n}, \ldots, r_{n-1:n} \) and all this mass is added to the weights at the extreme order statistics \( r_{1:n} \) and \( r_{n:n} \) in such a way that the barycenter (mean) remains the same.)

**Theorem 4.2.** \( D^c = (D_1, D_2, \ldots) \) is GC regular, and for every iid sequence \( x, D_n (x) \) convexly dominates \( F_n (x) \) for all \( n \). In particular, the means of \( F_n (x) \) and \( D_n (x) \) are the same, so if the underlying \( F \) has finite mean \( \mu \) then the mean of \( D_n (x) \) converges to \( \mu \) a.s.; also the variance of \( D_n (x) \) is at least the variance of \( F_n (x) \). In addition, if the support of \( F \) contains more than 2 points, this convex domination is eventually strict almost surely.
Proof. To see that $D^c$ is GC regular, note that $||D_n(x) - F_n(x)||_\infty \leq \max\{\delta_1, \delta_2\} \leq (n-2)(n \log n)^{-1}$, so (5) follows by the triangle inequality and Glivenko-Cantelli theorem.

It is easy to check (cf. [2]) that with the weights given in (13), $F_n(x)$ is a fusion of $D_n(x)$ for all $n$ and all iid $x$, so $F_n(x)$ is convexly dominated by $D_n(x)$. Since convex domination implies equal first moments, and higher second moments, this in turn implies that the variance of $D^c_n(x) \geq$ variance of $F_n(x)$ for all $n$ and $x$. The strictness conclusion follows since if the support of $F$ has at least 3 points, then eventually a.s. the support of $F_n(x)$ (and $D_n(x)$) will also have at least 3 points, in which case the fusion is nontrivial and the convex domination is strict.

To construct e.d.s.'s with smaller variance than the classical empirical distribution, simply reverse the process of $D^c$ by constructing fusions of $F_n$ in place of balayages.

Many other such constructions also produce convex domination or smaller variance; replacing the weights of 2.2(ii) (to form $D^+$) by non-monotone weights such as

$$w_{i:n} = n^{-1} + \left(-\left[i - \left(\frac{n+1}{2}\right)\right] + \frac{n^3 - n}{12}\right) g(n)$$

for appropriate small $g(n)$ produces GC regular e.d.s.'s which weight the outliers less than interior observations in a non-linear (in this case quadratic) fashion.

As one final example, suppose an estimator of $F$ is desired which will be as smooth as possible (in a given class). There are many ways to define the smoothness of a finitely-supported probability measure; the following is typical.

**Definition 4.3.** For real numbers $x_1 < x_2 < \cdots < x_m$ and increasing function $f : \mathbb{R} \to \mathbb{R}$ the smoothness $s$ of the graph of $f$ on $\{x_1, \ldots, x_m\}$ is

$$s = s(f; x_1, \ldots, x_m) = \max\{(f(x_j) - f(x_{j-1}))/(x_j - x_{j-1}) : 1 < j \leq n\}.$$

Similarly, for a finitely supported probability $G = \sum_{i=1}^n p_i \delta(y_i)$, $s(G)$ denotes $s(g; z_1, \ldots, z_m)$, where $z_1 < \cdots < z_m$ are the distinct values of $\{y_1, \ldots, y_n\}$, and $g(z_i) = G(y_j) \text{ if } z_i = y_j$. (Thus this smoothness is simply the maximum
slope of the linear interpolation of $f$ on $(x_1,x_m)$; note that a smaller value of $s$ denotes a smoother function.)

There are also many ways of smoothing a graph under given constraints; the following is one example.

**Definition 4.4.** For $\epsilon > 0$, $x_1 < \cdots < x_m$, and increasing function $f : \{x_1, \ldots, x_m\} \rightarrow \mathbb{R}^+$, let $\phi_\epsilon(f; x_1, \ldots, x_m)$ be the increasing positive function $g : \{x_1, \ldots, x_m\} \rightarrow \mathbb{R}^+$ which minimizes $s(h; x_1, \ldots, x_m)$ subject to

(i) $h : \{x_1, \ldots, x_m\} \rightarrow \mathbb{R}^+$ is increasing;
(ii) $\sum_{i=1}^m|h(x_i) - f(x_i)| = 0$;
(iii) $|h(x_i) - f(x_i)| \leq \epsilon$ for all $i = 1, \ldots, m$.

(Thus if $f$ represents a distribution function supported on $\{x_1, \ldots, x_m\}$, conditions (i) and (ii) guarantee that $g$ also will be; and condition (iii) implies that $\|f - g\|_{\infty} \leq \epsilon$. The existence of such a minimizing function $g$ is easy by compactness; constructions or algorithms for generating $\phi_\epsilon(f)$ are also easy to obtain.)

**Definition 4.5.** For $\epsilon > 0$, let

$$D_n^{s(\epsilon)}(r_1, r_2, \ldots) = \sum_{i=1}^m \phi_\epsilon(f; x_1, \ldots, x_m)\delta(x_i),$$

where $x_1 < x_2, \ldots < x_m$ are the distinct values of $\{r_1, \ldots, r_n\}$, and $f(x_i) = j/n$, where $j = \max\{k : r_k \leq x_i\}$; and for $\epsilon = (\epsilon_1, \epsilon_2, \ldots)$ a sequence of positive numbers, let $D_n^{s(\epsilon)} = (D_1^{s(\epsilon_1)}, D_2^{s(\epsilon_2)}, \ldots)$ be the corresponding e.d.s. (Informally, $D_n^{s(\epsilon)}$ is simply the $\phi_\epsilon$ smoothing of the classical distribution $F_n$.)

**Theorem 4.6.** For any sequence of positive numbers $\epsilon = (\epsilon_1, \epsilon_2, \ldots)$ converging to zero, $D_n^{s(\epsilon)}$ is GC regular, and for every iid sequence $X$, $s(D_n^{s(\epsilon)}(X)) \leq s(F_n(X))$ for all $n$ (i.e., $D_n^{s(\epsilon)}$ is smoother than $F_n$). Moreover this inequality is strict except when $F_n(X)$ is perfectly smooth ($s(F_n(X)) = 0$).

**Proof.** $D_n^{s(\epsilon)}$ is GC regular, since $\epsilon_n \rightarrow 0$ and since $\|D_n^{s(\epsilon_n)}(X) - F_n(X)\|_{\infty} \leq \epsilon_n$ for all $n$ and $X$. $D_n^{s(\epsilon_n)}(X)$ is smoother than $F_n(X)$ by construction (in fact it is the "$\epsilon_n$-smoothest"), and the strictness conclusion is obvious. \qed
ACKNOWLEDGEMENTS

The first author wishes to thank the Vrije Universiteit of Amsterdam, and Professor Piet Holewijn, for their hospitality during the summers of 1995 and 1996, and the Centro de Investigacion en Matematica (CIMAT) in Guanajuato and Professor Victor Perez-Abreu for their hospitality in early 1997. Both authors thank Pieter Allaart for several helpful suggestions.

REFERENCES