Response to Reviewer #1

First of all, I would like to thank this reviewer for his or her thoughtful comments and the potential simplifications he or she has pointed out.

A crucial statement can found towards the end of his / her report: "I don't understand why H_0 and H_1 are variables, and their expectations $\mu(H_0)$ and $\mu(H_1)$ and moments are mentioned." Therefore the referee thinks that, upon selection, X_i is a constant (being either zero or one).

However, the author treats the general case of two populations with nontrivial distributions and different expected values $\mu(H_0) \neq \mu(H_1)$. The assumption that $\mu(H_m) = m, (m = 0, 1)$ is only made in order to simplify the formulas. In particular, m, the index of the population, should not be confused with some realization $X_i = x_i$.

Despite this misconception, it is instructive to understand the main argument of the referee, i.e., to look at the trivial example of degenerate (constant) populations. Moreover, since the random variables X_i are organized in groups (subsamples), it is indeed helpful to consider $Z_j = \sum_{i=2^{j-1}}^{2^j-1} X_i$ $(j = 1, \dots, n)$ and the decomposition $S_n = \sum_{j=1}^n Z_j = \sum_{j=1}^n \sum_{i=2^{j-1}}^{2^j-1} X_i$.

However, it is not true that Z_j only assumes the values 2^{j-1} and 0 which would imply that $S_n \in \{0, \dots, 2^n - 1\}$, and S_n could be considered some "variant of binomial variable." Alas, since the X_i's have arbitrary distributions, $S_n = s_n$ may be any real number.

Rather, due to exponential sampling, the Weaver's distribution $W(n, p)$ extends $B(n, p)$ considerably. Since the basic building block of both distributions is the Bernoulli $B(p)$, they are related, but there is also a marked difference (see Definition 1 and Theorems 2-4). In particular, the paths defining the Weaver only split and never merge (see the Illustration, p. 8), and the limit distribution $W(p)$ is a fractal instead of being the Normal.

Nevertheless, since the selection process $\mathbf{B} = (B_0, \dots, B_{n-1})$ is a vector of independent Bernoullis, there is some "built-in" Binomial distribution. My proofs of Theorems 5, 6 and the first part of Theorem 11 rely on the construction process of $W(n, p)$, display the "built-in" cascade and/or weaving mechanism, and use the associated Binomial.

The referee is right that they can be simplified. The crucial observation is that the total variance in S_n (or, equivalently, in $\bar{X}_n$) stems from the independent Bernoulli selections (H_0 or H_1) and the variance in the distributions $\mathcal{L}(X_i)$. Since the Y_n's are *conditional expectations*, only the selection procedure is of relevance to them, and thus there is the elegant representation

$$(2^n - 1)y_{k,n} = E(S_n | (b_0, \dots, b_{n-1})) = \sum_{j=1}^n b_{j-1} 2^{j-1} \in \{0, \dots, 2^n - 1\},$$

mentioned in Theorem 2. This representation may be used in Theorem 5,

$$\begin{aligned} E(Y_n) &= E(\sum_{j=1}^n B_{j-1} 2^{j-1})/(2^n - 1)) = \frac{1}{(2^n - 1)} \sum_{j=1}^n E(B_{j-1} 2^{j-1}) \\ &= \frac{1}{(2^n - 1)} \sum_{j=1}^n 2^{j-1} E(B_{j-1}) = p, \end{aligned}$$

and in Theorem 6:

$$\begin{aligned}
\sigma^2(Y_n) &= \sigma^2(\sum_{j=1}^{n} B_{j-1}2^{j-1})/(2^n-1)) = \frac{1}{(2^n-1)^2}\sum_{j=1}^{n}\sigma^2(B_{j-1}2^{j-1}) \\
&= \frac{1}{(2^n-1)^2}\sum_{j=1}^{n}(2^{j-1})^2\sigma^2(B_{j-1}) = \frac{p(1-p)}{(2^n-1)^2}\sum_{j=1}^{n}2^{2(j-1)}
\end{aligned}$$

The expression $\sum_{j=1}^{n} 2^{2(j-1)}/(2^n-1)^2$ has a nice geometric interpretation that is used (implicitly) in the proof of Lemma 12.

	1	2	4	$\dots$	2^{n-1}	$\sum$
1	$\mathbf{1}$	2	4	$\dots$	2^{n-1}	2^n-1
2	2	$\mathbf{4}$	8	$\dots$	2^n	$2(2^n-1)$
4	4	8	$\mathbf{16}$	$\dots$	2^{n+1}	$4(2^n-1)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
2^{n-1}	2^{n-1}	2^n	2^{n+1}	$\dots$	$\mathbf{2^{2(n-1)}}$	$(2^{n-1})(2^n-1)$
$\sum$	2^n-1	$2(2^n-1)$	$4(2^n-1)$	$\dots$	$(2^{n-1})(2^n-1)$	$(2^n-1)^2$

That is, $\sum_{j=1}^{n} 2^{2(j-1)}$ is the trace of the above matrix and $\sum_{j=1}^{n} 2^{2(j-1)}/(2^n-1)^2$ is the proportion of the trace relative to the whole square.

For the first part of Theorem 11 note that $E\bar{X}_n = ES_n/(2^n-1)$, and

$$ES_n = E\left(\sum_{j=1}^{n}\sum_{i=2^{j-1}}^{2^j-1} X_i\right) = \sum_{j=1}^{n}\sum_{i=2^{j-1}}^{2^j-1} EX_i = \sum_{j=1}^{n}\sum_{i=2^{j-1}}^{2^j-1} p = (2^n-1)p,$$

since $EX_i = p\cdot\mu(H_1) + (1-p)\cdot\mu(H_0) = p$. (All random variables of some group j have distribution H_1 with probability p, and H_0 with probability $1-p$. Thus these are also the corresponding probabilities of any single X_i $(i = 1,\dots,2^n-1)$.)

The second part of Theorem 11 considers the variance $\sigma^2(\bar{X}_n)$. It turns out that it is crucial to study Z_j, which is the sum of 2^{j-1} random variables. Selection j chooses H_1 with probability p, and H_0 with probability $1-p$. Given H_m $(m = 0,1)$, the conditional variance of Z_j is $\sigma^2(Z_j|H_m) = 2^{j-1}\sigma^2(X_i|H_m) = 2^{j-1}\sigma_m^2$, since the X_i's are independent.

Moreover, $E(Z_j|H_m) = 2^{j-1}m$, and $EZ_j = 2^{j-1}p$. A variance decomposition of Z_j yields

$$\begin{aligned}
\sigma^2(Z_j) &= p(E(Z_j|H_1)-EZ_j)^2 + (1-p)(E(Z_j|H_0)-EZ_j)^2 \\
&\quad + p\sigma^2(Z_j|H_1) + (1-p)\sigma^2(Z_j|H_0) \\
&= p(2^{j-1}-2^{j-1}p)^2 + (1-p)(0-2^{j-1}p)^2 + p2^{j-1}\sigma_1^2 + (1-p)2^{j-1}\sigma_0^2 \\
&= 2^{2(j-1)}p(1-p)(1-p+p) + 2^{j-1}p\sigma_1^2 + 2^{j-1}(1-p)\sigma_0^2
\end{aligned}$$

which implies

$$\begin{aligned}\sigma^2(\bar{X}_n) &= \sigma^2\left(\frac{S_n}{2^n-1}\right) = \frac{1}{(2^n-1)^2}\sigma^2(S_n) = \frac{1}{(2^n-1)^2}\sum_{j=1}^{n}\sigma^2(Z_j) \\ &= \frac{\sum_{j=1}^n 2^{2(j-1)}}{(2^n-1)^2}p(1-p) + \frac{\sum_{j=1}^n 2^{j-1}}{(2^n-1)^2}p\sigma_1^2 + \frac{\sum_{j=1}^n 2^{j-1}}{(2^n-1)^2}(1-p)\sigma_0^2 \\ &= \sigma^2(Y_n) + \frac{p\sigma_1^2 + (1-p)\sigma_2^2}{2^n-1}. \end{aligned} \tag{1}$$

Consistently, the variance within the population washes out quickly, and one obtains $\lim_{n\to\infty}\sigma^2(\bar{X}_n) = \sigma^2(Y) = p(1-p)/3$. This means, that the limit distribution of $\bar{X}_n$ is the same as that of Y_n, i.e., $W(p)$. In other words, within the populations, the LLN applies. Therefore, in the limit, only the variance caused by exponential sampling (the selection process) is relevant.

Although equation (7) is wrong and should be replaced by the above equation (1), the idea of Theorem 13 is still valid, and thus that Theorem may also be "repaired:"

Starting with $Y_n \sim W(n,p)$, one may replace each realization $y_{k,n} = k/(2^n-1)$ by a Bernoulli r.v. $C_{k,n} \sim B(y_{n,k})$. The r.v. $C_{k,n} \circ Y_n$ assumes the values zero and one, since each $y_{k,n}$ is mapped to 1 w.p. $y_{k,n}$ and to 0 w.p. $1-y_{k,n}$. Since the location of the distribution does not change if on splits the realizations $y_{k,n}$ in the aforementioned way, $E(C_{k,n} \circ Y_n) = EY_n = p$, and thus $C_{k,n} \circ Y_n \sim B(p)$. For fixed n, one might think of the collection of all $C_{k,n}$ ($k = 0,\ldots,2^n-1$) as a family of "dual distributions" to $W(k,n)$. In the paper, this train of thought is called "mixing," and it is still true that the total variance $p(1-p)$ is the sum of "weaving and mixing."

In a nutshell, owing to the "elementary" construction principles used throughout the manuscript, explicit formulas can be given for the crucial parameters of the processes and their limits. Because of Theorem 10, at least some of these results can be applied to the received situation. Moreover, Theorem 10 demonstrates that a multifractal discrete structure (i.e., $W(p)$) derived in an iterative "bottom up" way[1] may be equivalent to a structure obtained with a similar iterative "top down" procedure.[2] At least to the author, this looks like an exemplar of a more general sandwich principle.

Of course, to some extent, it is a matter of taste how many of the extensions should be discussed in Section 7.

Finally, the author agrees with the reviewer that the core of this contribution is a mathematical analysis of an important nonlinear model that is applied in many fields.

[1] starting with $B(p)$ or point mass at p, if one defines $W(k,0) = \varepsilon_p$

[2] starting with the uniform distribution on the unit interval