# A note on demi-supermartingales for sampling without replacement

Chase Mathis
Department of Statistics
University of California, Berkeley

cmathis@berkeley.edu

August 25, 2026

**Abstract**

We study the problem of constructing confidence intervals for the mean of a finite population after sampling without replacement. We find that the Symmetric Polynomial (SymPol) interval introduced by Ming et al. [2026] is valid not only for so-called "co-valid" e-variables, but also for negatively dependent e-variables constructed under a without-replacement model. Gaffke's interval, introduced by Gaffke [2005] and proved by Vlassis and Thomas [2026], is not valid under this model. Moreover, we show that symmetric polynomials built from a sequential e-process constructed under the without-replacement model are valid and enjoy several other desirable properties. Simulations demonstrate that, among a variety of intervals for sampling without replacement, the modified SymPol interval is the shortest.

## 1 Introduction

Gaffke [2] gave two conjectures relating to inference on the mean of independent bounded random variables. Recently, Ming et al. [6] resolved one conjecture. Vlassis and Thomas [12] resolved the other conjecture. We state the recently proven conjectures now.

**Theorem 1.1** (Gaffke's interval for bounded means)**.** *Let $(X_i)_{i=1}^n$ be nonnegative independent random variables such that $\forall i \in [n], \mathbb{E}_{\mathsf{P}}[X_i] = 1$. Gaffke first conjectured that*

$$\mathsf{P}\left(\sup_{\lambda\in[0,1]} \prod_{i=1}^{n} (1 - \lambda + \lambda X_i) \geqslant 1/\alpha\right) \leqslant \alpha. \tag{1}$$

*Gaffke's second conjecture is*

$$\mathsf{P}(K_n(X_1, \ldots, X_n) \leqslant \alpha) \leqslant \alpha \quad \text{where} \quad K_n(\mathbf{x}) := \mathsf{P}_{\mathbf{D}}\left(\sum_{i=1}^{n} x_i D_i \leqslant 1\right), \tag{2}$$

*where $(D_0, D_1, \ldots, D_n) \sim \mathrm{Dirichlet}(1, \ldots, 1)$. We note that (1) was first stated by Wang and Zhao [14] under the stronger assumption that the data are i.i.d..*

Readers familiar with the e-value literature [9, 13] may notice that these statements may be abstractly thought of as e-to-p mergers. Beyond the theoretical interest, the conjectures are of practical use in that they enable the analyst to construct sharp nonasymptotic confidence intervals for the mean of independent random variables whose bounds are known, improving upon older results in the literature from Hoeffding [3] and Bernstein. Additionally, Ming et al. [7] showed that the interval constructed from (2) is inadmissible but asymptotically efficient.

In proving (1), Ming et al. [6] explored a new class of "Symmetric Polynomial" (SymPol) confidence intervals and used demi-supermartingales [8] to prove the results. We will refer to (1) as the "regret-free" interval and the SymPol interval which dominates it. We will refer to (2) simply as Gaffke's interval.

While Theorem 1.1 is stated assuming the data are independent, Ming et al. [6] demonstrate that the regret-free and SymPol intervals are valid under a certain type of condition called "co-validity", which is strictly weaker than independence. In addition, Vlassis and Thomas [12] show that Gaffke's interval is valid if the data are independent but not identically distributed, but the variables must share a common mean.

The present paper asks two questions. *First, do (1) and (2) remain valid under a model of sampling without replacement? Second, can one modify either expression to obtain sharp intervals that shrink as $n \to N$, as Serfling-type [10, 1] intervals do?* We state the answers in the following brief and informal theorem.

**Theorem 1.2** (SymPol intervals for sampling without replacement; *brief & informal*)**.** *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of nonnegative numbers in $[0,1]$ with mean $m$. Then sample $n \leqslant N$ values from this population without replacement to obtain the random vector*

$$\mathbf{X} \equiv (X_1, \ldots, X_n)\,.$$

*The SymPol interval remains valid under this model, but Gaffke's interval is invalid. Mathematically, define*

$$\widetilde{E}_i = X_i/m \quad \text{and} \quad E_i \equiv \frac{X_i}{\mu_i^{\mathrm{WoR}}} \quad \text{where} \quad \mu_i^{\mathrm{WoR}} = \frac{Nm - \sum_{j=1}^{i-1} X_i}{N-i+1}.$$

*Then*

$$\mathsf{P}\left(\sup_{\lambda\in[0,1]}\prod_{i=1}^{n}(1-\lambda+\lambda\widetilde{E}_i) \geqslant 1/\alpha\right) \leqslant \alpha \quad \text{and} \quad \mathsf{P}\left(\sup_{\lambda\in[0,1]}\prod_{i=1}^{n}(1-\lambda+\lambda E_i) \geqslant 1/\alpha\right) \leqslant \alpha, \tag{3}$$

*but there exist populations $\mathcal{X}$, sizes $n$, and levels $\alpha$ for which*

$$\mathsf{P}\Big(K_n(\widetilde{E}_1, \ldots, \widetilde{E}_n) \leqslant \alpha\Big) > \alpha.$$

There are various surprising aspects contained in Theorem 1.2. First, notice that $\widetilde{E}_i$ are neither independent nor co-valid, but satisfy (3). Co-validity is a central part of the proof techniques in [6], but this work demonstrates that a wider class of $e$-values satisfy the demi-supermartingale reduction. In addition, Gaffke's interval lacks validity under sampling without replacement. Concentration inequalities derived from Chernoff's method that are valid for independent data hold for data sampled without replacement. The result allows the concentration inequalities from Hoeffding [3], Bernstein, and Massart [5] to be applied to data sampled without replacement. Lastly, the theorem claims that we can sharpen the SymPol confidence intervals by explicitly using the without replacement structure. These intervals enjoy all the benefits of the regular SymPol interval, but also shrink as $n \to N$.

In Section 2, we will make precise the problem setup of sampling without replacement and recount results that are valid and exploit the without-replacement structure. In Section 3, we expand upon each statement in Theorem 1.2 and demonstrate through simulations the advantage of the SymPol interval over other methods.

## 2 Sampling without replacement

In this section, we recall the problem of inference in the setting of sampling without replacement (WoR). Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a finite population of fixed points. Importantly, the $N$ values in the population are fixed and considered nonrandom. The randomness is derived from the order in which the numbers are revealed to the analyst. A simple model describing this data generation is

$$X_t \mid (X_1, \ldots, X_{t-1}) \sim \text{Uniform}\left((x_1, \ldots, x_N) \backslash (X_1, \ldots, X_{t-1})\right).$$

Another useful model is to define

$$\Sigma_N = \text{set of all permutations on } [N],$$

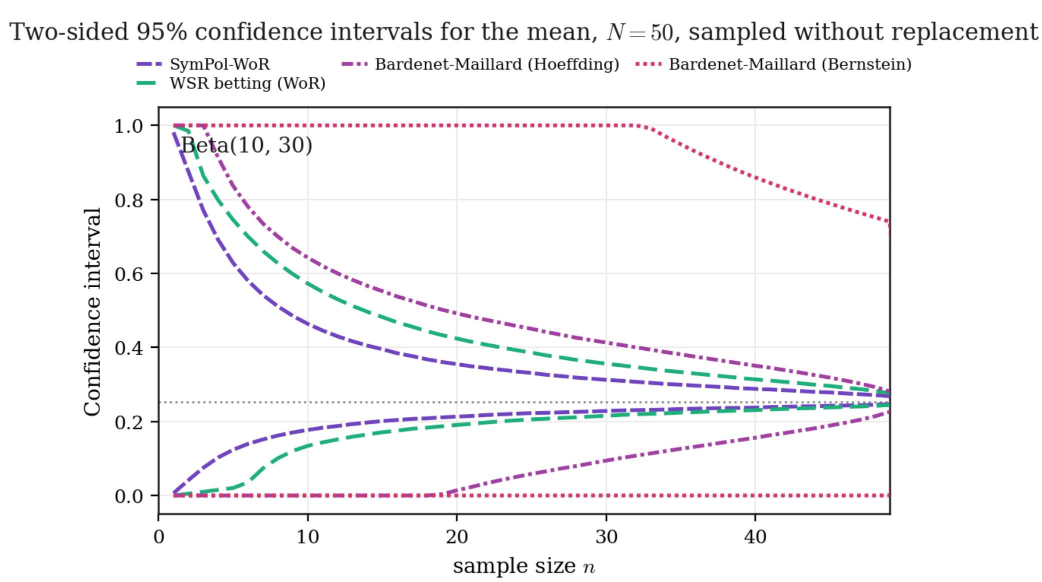
draw a permutation $\pi$ uniformly at random from $\Sigma_N$, then set $X_i = x_{\pi(i)}$. We will appeal to both models in this work.

For the present paper, the parameter of interest is the mean $\mu \equiv N^{-1} \sum_{i=1}^N x_i$. We wish to construct a $(1-\alpha)$-confidence interval $C_n \equiv C_n(X_1, \ldots, X_n)$ such that

$$\mathsf{P}(\mu \in C_n) \geqslant 1 - \alpha. \tag{4}$$

Note the probability statement in (4) is required to hold in finite samples and thus large-sample confidence intervals supported by finite-population CLTs [4] are not considered. We will assume that $0 \leqslant \min \mathcal{X} \leqslant \max \mathcal{X} \leqslant 1$ without loss of generality for bounded data with known bounds.[1] We now recount the established methods one can use to compute $C_n$.

[1] If the analyst instead knows the data lie in $[a, b]$, they can rescale the data to lie in $[0, 1]$.



## 2.1 Existing confidence intervals

It is oftentimes noted that random variables from sampling without replacement concentrate *more* tightly than i.i.d. draws. The statement builds on [3, Theorem 4], which allows Chernoff-type concentration inequalities to be applied to WoR draws. Thus if the analyst knows that $0 \leqslant \min \mathcal{X} \leqslant \max \mathcal{X} \leqslant 1$, Hoeffding's interval

$$C_n^{\mathrm{H}} \equiv \widehat{\mu}_n \pm \sqrt{\frac{\log(2/\alpha)}{2n}}$$

satisfies (4).

With this familiar confidence interval in mind, we explain why we want to continue a search for a "better" confidence interval than Hoeffding's. We define "better" in the following four desiderata. We want $C_n$ to be (1) valid in finite samples, (2) adapt to the variance of the distribution, (3) adapt to the asymmetry of the distribution, and (4) shrink as $n \to N$.[2]

Hoeffding's inequality satisfies the first desideratum, but fails in the latter three. While Bernstein's confidence interval [11] $C_n^{\mathrm{B}}$ is nonasymptotic and variance-adaptive, it fails to adapt to the structure of sampling without replacement and is symmetric. Serfling [10] improves upon the inequalities from Hoeffding and Bernstein by explicitly using the WoR model. These inequalities are further refined by Bardenet and Maillard [1]. We give the Bardenet–Maillard refined Serfling bounds in the following proposition.[3]

**Proposition 2.1** (Hoeffding– and Empirical Bernstein–Serfling Intervals [1, Corollary 2.5, Theorem 4.3])**.** *Let $X_1, \ldots, X_n$ be $n$ samples without replacement from $\mathcal{X}$ where $0 \leqslant \min \mathcal{X} \leqslant \max \mathcal{X} \leqslant 1$. Then the Hoeffding–Serfling interval satisfies, with probability at least $1-\alpha$,*

$$\mu \in \hat{\mu}_n \pm \sqrt{\frac{\rho_n \log(1/\alpha)}{2n}} \quad \text{where } \rho_n = \begin{cases} 1-(n-1)/N & n \leqslant N/2 \\ (1-n/N)(1+1/n) & n > N/2 \end{cases}. \tag{5}$$

*The Empirical Bernstein–Serfling interval satisfies, with probability at least $1-5\alpha$,*

$$\mu \in \hat{\mu}_n \pm \hat{\sigma}_n \sqrt{\frac{2\rho_n \log(2/\alpha)}{n}} + \frac{\kappa \log(2/\alpha)}{n}, \tag{6}$$

*where $\rho_n$ is the same as above and $\kappa = 7/3 + 3/\sqrt{2}$.*

Notice the Hoeffding–Serfling confidence interval in (5) uses the sampling without replacement structure, shrinking as $n$ approaches $N$. When $n = N$, the confidence interval is a single point. The Bernstein–Serfling confidence interval in (6) satisfies the desired scaling, in addition to being adaptive to the variance. Nevertheless, the Bernstein–Serfling interval is symmetric around the mean and thus *does not* adapt to the asymmetry of the data distribution.

The last interval we recount is from Waudby-Smith and Ramdas [15]. The authors develop a "betting" framework that satisfies all four desiderata; it is the main result the present work builds upon. While the bulk of the authors' attention was on building time-uniform confidence *sequences* for i.i.d. data, there is some attention to pointwise confidence *intervals* under sampling without replacement.

[2] For the last desideratum, we note that when $n = N$ the analyst has observed the full population and there is no randomness left.

[3] The sampling without replacement bounds are "time-uniform"–the probability statements are valid for all $k \in [n]$ and hence can be used for sequential analysis. We only consider the pointwise valid bounds (see [1, Corollary 2.5] and [1, Corollary 4.3]).

Waudby-Smith and Ramdas [15] define the following WoR "capital process" for a mean $m$

$$\mathcal{K}_n^{\mathrm{WoR}}(\mathbf{X}; m) := \prod_{i=1}^{n} \left(1 + \gamma_i(m)(X_i - m_i^{\mathrm{WoR}})\right) \quad \text{where} \quad m_i^{\mathrm{WoR}} := \frac{Nm - \sum_{j=1}^{i-1} X_i}{N - (i-1)}. \tag{7}$$

One can rearrange this process to be "normalized" so that

$$\mathcal{K}_n(\mathbf{X}; m) := \prod_{i=1}^{n} (1 - \lambda_i + \lambda_i E_i) \quad \text{where} \quad E_i \equiv \frac{X_i}{m_i^{\mathrm{WoR}}}.$$

In (7), we must require $\gamma_i(m) \in [-1/(1 - m_i^{\mathrm{WoR}}), 1/m_i^{\mathrm{WoR}}]$, but the normalized process allows $\lambda_i \in [0,1]$. If $(\lambda_i)_{i=1}^n$ are chosen predictably, then (7) is an e-process and it can be inverted to form a confidence interval. Waudby-Smith and Ramdas [15] discuss different algorithms to choose the "betting strategy", under the constraint that $\lambda_i$ can only depend on $X_1, \ldots, X_{i-1}$. Using different heuristic online algorithms, the confidence intervals derived from (7) satisfy all four desiderata and dominate (empirically) the aforementioned confidence intervals from [10, 1].

While this interval indeed satisfies the four desiderata, one peculiarity is the lack of symmetry with respect to the vector $\mathbf{X}$. One may want the confidence interval generated from (7) be the same regardless of the order of the data. While the data are not independent, they are still exchangeable, and therefore the order of coordinates should not change the interval. We will discuss this further when comparing the SymPol interval and the SymPol-WoR interval.

## 3 SymPol and Gaffke under a without-replacement model

In this section, we describe the novel SymPol-WoR interval that is valid and empirically dominates various other intervals. In addition, we will show that the naïve SymPol interval applied to the without-replacement e-variables remains valid, but the naïve Gaffke interval is not always valid.

### 3.1 The SymPol-WoR interval

In order to construct confidence intervals using the capital processes in (7), one must specify an online algorithm that gives $(\lambda_i)_{i=1}^n$. Waudby-Smith and Ramdas [15] give various "betting strategies" aimed at minimizing the regret:

$$\mathcal{R}_n = \sup_{\lambda \in [0,1]} \prod_{i=1}^{n} (1 - \lambda + \lambda E_i) - \prod_{i=1}^{n} (1 - \lambda_i + \lambda_i E_i).$$

While one would prefer to use the hindsight-optimal capital process (the term on the left), we are constrained so that we can only adapt the betting strategy based on the observations we've seen. The first theorem of this paper shows that the hindsight-optimal, regret-free capital process satisfies Markov's inequality and thus can be inverted to form a confidence interval.

**Theorem 3.1** (Regret-free confidence intervals for sampling WoR)**.** *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a population of $N$ fixed real numbers in $[0,1]$ and $X_1, \ldots, X_n$ a sample without replacement. Then for $E_i = X_i/m_i^{\mathrm{WoR}}$,*

$$\mathsf{P}\left(\sup_{\lambda \in [0,1]} \prod_{i=1}^{n} (1 - \lambda + \lambda E_i) \geqslant 1/\alpha\right) \leqslant \alpha. \tag{8}$$

*This interval is pointwise dominated by the SymPol-WoR interval where for*

$$B_k(E_1, \ldots, E_n) = \frac{\sum_{|I|=k, I \subset [n]} \prod_{i \in I} E_i}{\binom{n}{k}}, \quad \mathsf{P}\left(\max_{0 \leqslant k \leqslant n} B_k(E_1, \ldots, E_n) \geqslant 1/\alpha\right) \leqslant \alpha. \tag{9}$$

The proof of Theorem 3.1 is in Section 4. It is constructed by extending the demi-supermartingale theory developed by Ming et al. [6] and hence directly proves the pointwise sharper SymPol-WoR confidence interval in (9). For many distributions, the intervals constructed by (8) and (9) are indistinguishable. However, for data distributions that have many zeros, the SymPol-WoR is more stable than the regret-free interval (see Fig. 1).

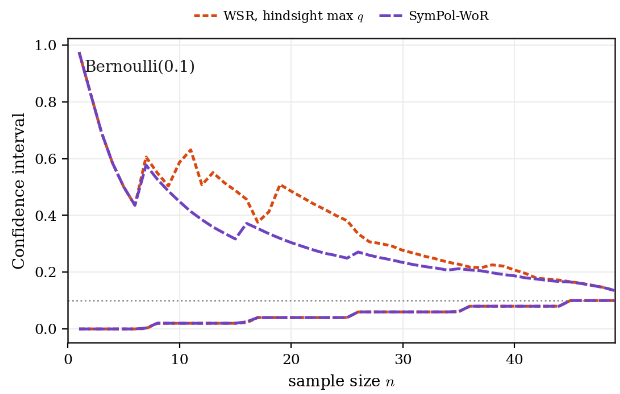



Figure 1: The stability of the SymPol-WoR interval more stable than the regret-free interval.

It is not immediate that the regret-free confidence interval or the SymPol-WoR interval dominates the predictable plug-in interval. While regret is constrained to use the single best $\lambda$, online algorithms can vary $\lambda_i$. Hence, the online algorithm has access to a richer set of functions and thus may be sharper than the singular best $\lambda$ in hindsight. Despite this possibility, simulations we consider suggest that the SymPol-WoR bound is as tight as or tighter than the predictable algorithms given by Waudby-Smith and Ramdas [15] (see Fig. 2).

### 3.2 Validity of naïve SymPol

We now show that using the default SymPol interval, which was proved for co-valid $e$-variables in [6], is also valid for the negatively dependent $e$-variables generated by the WoR model.

Suppose an analyst has learned that sampling without replacement concentrates *more* tightly than taking independent draws. They decide to use the default SymPol interval for inference in the WoR model. They define

$$\widetilde{E}_i = X_i/m$$

where $m = \sum_{i=1}^{N} x_i/N$ is assumed under the null hypothesis. The question is whether

$$\mathsf{P}\left(\sup_{0\leqslant k\leqslant n} B_k\left(\widetilde{E}_1,\ldots,\widetilde{E}_n\right) \geqslant 1/\alpha\right) \leqslant \alpha. \tag{10}$$

While Ming et al. [6] proved (10) if $(\widetilde{E}_i)_{i=1}^n$ are co-valid, the present model assumes that the $e$-variables are negatively dependent. Notice that this dependence is strictly different from the dependence we prove in Theorem 3.1. One way to see the difference is to note that $(\widetilde{E}_i)_{i=1}^n$ are now exchangeable whereas $(E_i)_{i=1}^n$ are not.

**Proposition 3.1** (SymPol is valid for WoR $e$-variables)**.** *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a population of $N$ fixed real numbers in $[0, 1]$ and $X_1, \ldots, X_n$ a sample without replacement from $\mathcal{X}$. Then for*

$\widetilde{E}_i = X_i/m$ *under the null that the mean of the full population is* $m$,

$$\mathsf{P}\left(\max_{0\leqslant k\leqslant n} B_k(\widetilde{E}_1,\ldots,\widetilde{E}_n) \geqslant 1/\alpha\right) \leqslant \alpha.$$

*This implies the regret-free interval*

$$\mathsf{P}\left(\sup_{\lambda\in[0,1]} \prod_{i=1}^{n} \left(1-\lambda+\lambda\widetilde{E}_i\right) \geqslant 1/\alpha\right) \leqslant \alpha.$$

The proof of Proposition 3.1 is in Section 5. The result provides the SymPol interval with another class of valid e-variables strictly beyond co-validity and independence. The intervals described in Proposition 3.1 have some benefits and drawbacks compared with the ones described in Theorem 3.1. One drawback is that the intervals in Proposition 3.1 do not shrink as $n \to N$, eventually becoming degenerate when $n = N$. However, one benefit of these intervals is that they are agnostic to the ordering, respecting the exchangeability of the data. Therefore, reconsidering Section 2, we may search for an interval that satisfies all four desiderata and in addition respects the exchangeability of the data. It is unclear if this is possible.

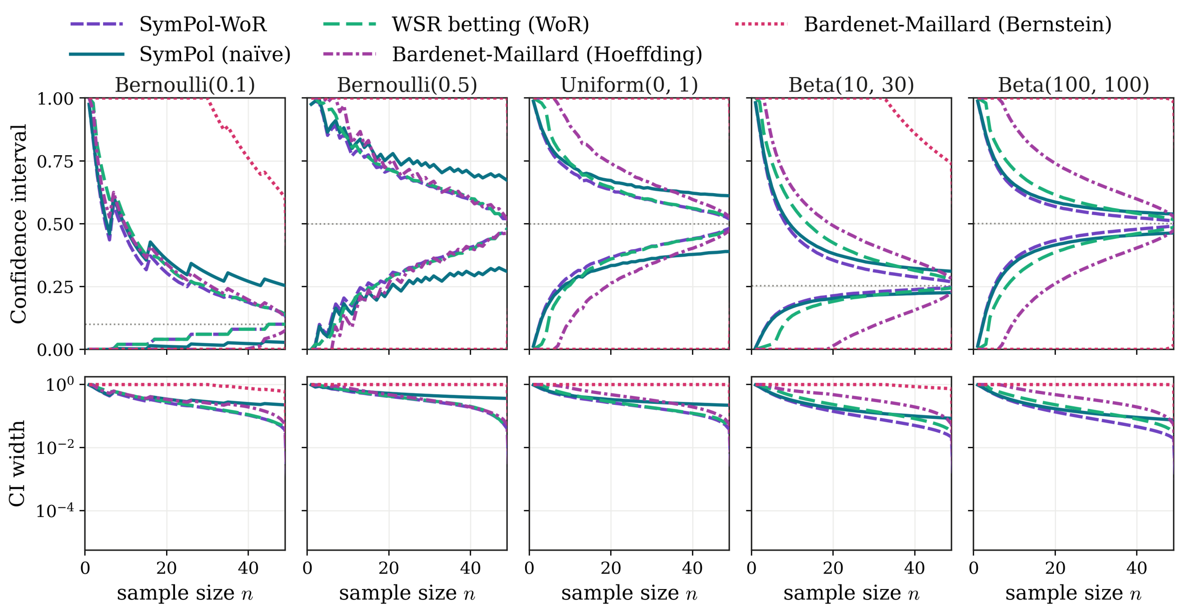



Figure 2: Confidence intervals for the mean under different distributions. Classical Serfling intervals [1] are wider than the betting-style intervals [15]. The SymPol-WoR interval dominates the interval based on the predictable plug-in $\lambda_n$.

Finally, naïve SymPol is not valid for all negatively dependent random variables: the proof relies on the without-replacement structure. Negative dependence is a definition quite useful when computing concentration inequalities using MGFs. In these proofs, the main bound is an expectation of products. The SymPol and Gaffke intervals do not use this structure and operate on different functions.

### 3.3 Invalidity of naïve Gaffke

In the previous section, we showed that naïvely using the SymPol interval for WoR inference does no harm—the interval remains valid. Here, we show that this is not true of Gaffke's interval. Gaffke's interval relies on independence in a way that sampling without replacement may fail in certain cases. We repeat that this result is surprising given the longstanding belief that sampling without replacement concentrates more than independent sampling. The result may also shed light on the peculiarities of what makes Gaffke's interval so remarkably tight.

**Theorem 3.2** (Gaffke's interval is invalid for WoR $e$-variables)**.** *Let $\mathcal{X} = (x_1, \ldots, x_N)$ be a population of $N$ fixed real numbers in $[0,1]$ and $X_1, \ldots, X_n$ be a sample without replacement from $\mathcal{X}$. With $\widetilde{E}_i = X_i/m$, there exist populations $\mathcal{X}$ and levels $\alpha$ for which Gaffke's interval is invalid:*

$$\mathsf{P}\Big(K_n(\widetilde{E}_1, \ldots, \widetilde{E}_n) \leqslant \alpha\Big) > \alpha. \tag{11}$$

*Proof of Theorem 3.2.* Let $\mathcal{X} = (0,0,1)$ and draw $n = 2$ samples from $\mathcal{X}$. One can calculate that

$$\mathsf{P}\Big(K_n(\widetilde{E}_1, \widetilde{E}_2) \leqslant 5/9\Big) = 2/3 > 5/9.$$

A counterexample that does not have a spike structure is $\mathcal{X} = (1/12, 11/12, 1)$, $n = 2$, and $\alpha \geqslant 101/110$. □

Although (11) shows that Gaffke cannot be applied to $\widetilde{E}_1, \ldots, \widetilde{E}_n$, we do not claim that applying Gaffke to the sequential $e$-process used in Theorem 3.1 is invalid. It may be interesting to explore whether $K_n(E_1, \ldots, E_n)$ is a valid $p$-value under sampling without replacement. We are unable to prove or disprove the super-uniformity of $K_n(E_1, \ldots, E_n)$.

Some evidence suggests that even if $K_n(E_1, \ldots, E_n)$ is super-uniform, it may not be the right object of study for sampling WoR. One can show that $K_n(E_1, \ldots, E_n)$ is not first-order asymptotically efficient. Ming et al. [7] show that Gaffke's interval for i.i.d. data *is* asymptotically efficient. Further study is needed to identify the correct analogue of Gaffke's interval when the data are sampled without replacement.

# 4 Proof of SymPol-WoR

Before we begin, we recall the notation. Let $\mathcal{X} = (x_1, \ldots, x_N) \in [0,1]^N$. We sample $X_1, \ldots, X_n$ without replacement by sampling $\pi \sim \Sigma_N$ the set of all permutations on $[N]$ and setting $X_i = x_{\pi(i)}$.

Before proceeding with the proof, we begin by putting an ordering on the random orderings $\pi$. The random objects for sampling without replacement are the orderings. Therefore, we can have a similar notion of "decreasing" when the randomness stems from orderings rather than values.

**Step 0: A partial order on $\Sigma_N$** Let $\pi$ be an ordering of $[N]$. Let $s_t$ be a function of $\pi$ that swaps the values in positions $t$ and $t+1$. We say $s_t$ is an *up-move* if it places the larger of two items into the earlier position. For instance, let the random realization of $\{1,2,5\}$ be $(1,5,2)$. We have that $s_1$ is an up-move while $s_2$ is not an up-move. The partial order we give to the random orderings is as follows.[4] We say $\pi \preceq \pi'$ if $\pi'$ is obtainable by a composition of up-moves. For example,

$$(2,4,6) \preceq (2,6,4) \preceq (6,2,4) \preceq (6,4,2).$$

We say a function $G$ from orderings to reals is *swap-decreasing (increasing)* if when $\pi \preceq \pi'$ then $G(\pi') \overset{(\geqslant)}{\leqslant} G(\pi)$. Equivalently, $G(\pi)$ is swap-decreasing if when $s_i$ is an up-move then $G(s_i \circ \pi) \leqslant G(\pi)$.

[4] Note that this ordering is partial. We cannot compare $(4,2,6)$ with $(6,2,4)$.

The central functions we will work with are the symmetric polynomials, which we recall below:

$$S_k := S_k(E_1, \ldots, E_n) = \sum_{|I|=k, I \subset [n]} \prod_{i \in I} E_i, \quad \text{and} \quad B_k := \frac{S_k}{\binom{n}{k}}.$$

The randomness stems from the ordering $\pi$, so it is more convenient to think about $B_k$ as functions from the orderings to the reals. By Lemma 4.5 the symmetric polynomials are swap-increasing.

**Notation.** Let

$$T_j(\pi) = \sum_{i=j}^{N} x_{\pi(i)}, \quad r_j = N - j + 1, \quad m_j(\pi) = \frac{T_j(\pi)}{r_j} \quad \text{and} \quad E_j(\pi) = \frac{x_{\pi(j)}}{m_j(\pi)},$$

where we write $T = T_1, m = m_1$ and note $N = r_1$. If $T_i(\pi) = 0, \forall i \geqslant j$ then as a convention set $E_i = 1$ for $i \in \{j, \ldots, N\}$. For $S \subset [n]$, write $E_S = \prod_{i \in S} E_i$ (and $E_\varnothing = 1$).

Notice that if $\pi(i) = \pi'(i)$ for all $i = 1, \ldots, j-1$ then $T_j(\pi) = T_j(\pi')$ (and hence $m_j(\pi) = m_j(\pi')$). If a function $W(\pi)$ satisfies this property we will say that $W(\pi)$ is $j-1$-invariant. A corollary of a function being $j-1$-invariant is that if $i > j-1$ then a swap at position $i$ keeps the function the same, meaning $W(s_i \circ \pi) = W(\pi), \forall i \in \{j, \ldots, N\}$.

**Lemma 4.1.** *Let $j \in [N-1]$. Suppose $F(\pi) \geqslant 0$, and $F(s_j \circ \pi) = F(\pi)$. Suppose $G(\pi) \geqslant 0$ is swap-decreasing. Then,*

$$\sum_{\pi \in \Sigma_N} \left(x_{\pi(j+1)} - x_{\pi(j)}\right) F(\pi) G(\pi) \geqslant 0$$

*Proof of Lemma 4.1.* Fix $j \in [N-1]$ and call $x_{\pi(j+1)} - x_{\pi(j)} = D(\pi)$. Notice that $D(s_j \circ \pi) = -D(\pi)$. Moreover, $s_j \circ \Sigma_N = \Sigma_N$. Therefore

$$\begin{aligned}\sum_{\pi \in \Sigma_N} D(\pi) F(\pi) G(\pi) &= \sum_{\pi \in \Sigma_N} D(s_j \circ \pi) F(s_j \circ \pi) G(s_j \circ \pi) \\ &= -\sum_{\pi \in \Sigma_N} D(\pi) F(\pi) G(s_j \circ \pi).\end{aligned}$$

We can add the two equations to get

$$2 \sum_{\pi \in \Sigma_N} D(\pi) F(\pi) G(\pi) = \sum_{\pi \in \Sigma_N} D(\pi) F(\pi) \left(G(\pi) - G(s_j \circ \pi)\right) \tag{12}$$

We will focus on the summands in (12).

**Case 1:** $x_{\pi(j+1)} - x_{\pi(j)} \leqslant 0$ In this first case we have $D(\pi) \leqslant 0$. Because $G$ is swap-decreasing, it follows that $G(\pi) \leqslant G(s_j \circ \pi)$ and hence their difference is $\leqslant 0$. Therefore each summand is nonnegative.

**Case 2:** $x_{\pi(j+1)} - x_{\pi(j)} > 0$ Here, $D(\pi) \geqslant 0$ and so is $G(\pi) - G(s_j \circ \pi)$. Therefore each summand is nonnegative. Since each summand is nonnegative the sum is also nonnegative. $\square$

Now, we state the main lemma.

**Lemma 4.2.** *Let $1 \leqslant j \leqslant N$ and $S \subset \{j+1, \ldots N\}$. Let $G(\pi) \geqslant 0$ swap-decreasing and $W(\pi) \geqslant 0$ be j-1-invariant. Then*

$$\sum_{\pi \in \Sigma_N} W(\pi)\left(m_j - x_{\pi(j)}\right) E_S(\pi) G(\pi) \geqslant 0.$$

*Proof of Lemma 4.2.* We prove by reverse induction on $j$. Let the base case be $j = N$. Then $S = \varnothing$ and $m_j = x_{\pi(j)}$ so each summand is exactly zero. We prove the induction step at index $j$. Thus assume for all indices greater than $j$, the claim holds. Set $a = \min S$. When $S = \varnothing$, we can have $a = N + 1$.

**Case 1:** The first case assumes that $j + 1 \notin S$ and thus $a - 1 > j$. We first use the decomposition from Lemma 4.6

$$\begin{aligned} r_j(m_j - x_{\pi(j)}) &= \sum_{i=j}^{N-1} (N-i)(x_{\pi(i+1)} - x_{\pi(i)}) \\ &= \sum_{i=j}^{a-2} (N-i)(x_{\pi(i+1)} - x_{\pi(i)}) + \sum_{i=a-1}^{N-1} (N-i)(x_{\pi(i+1)} - x_{\pi(i)}) \\ &= \underbrace{\sum_{i=j}^{a-2} (N-i)(x_{\pi(i+1)} - x_{\pi(i)})}_{\mathrm{I}} + \underbrace{r_{a-1}(m_{a-1}(\pi) - x_{\pi(a-1)})}_{\mathrm{II}}. \end{aligned}$$

Multiply both sides by $W(\pi)E_S(\pi)G(\pi)$ and then take summations over $\pi \in \Sigma_N$ on both sides. With this we obtain for the first term I,

$$\sum_{i=j}^{a-2} (N-i) \sum_{\pi \in \Sigma_N} W(\pi)\left(x_{\pi(i+1)} - x_{\pi(i)}\right) E_S(\pi) G(\pi).$$

For every $j \leqslant i \leqslant a-2$, the $j-1$-invariance of $W$ gives $W(s_i \circ \pi) = W(\pi)$. The same swap leaves $E_S$ unchanged because all indices in $S$ are at least $a$. With this observation, call the product $W(\pi)E_S(\pi) = F(\pi)$, since $F(s_i \circ \pi) = F(\pi)$ and by assumption $G(\pi)$ is swap-decreasing. Then apply Lemma 4.1 term-wise to see that term I is nonnegative.

For the second term multiply $W(\pi)E_S(\pi)G(\pi)$ and sum over $\pi \in \Sigma_N$ to obtain

$$r_{a-1} \sum_{\pi \in \Sigma_N} \left\{ W(\pi)(m_{a-1}(\pi) - x_{\pi(a-1)}) E_S(\pi) G(\pi) \right\}.$$

Notice that $a - 1 > j$ and $S \subset \{a, \ldots, N\}$. Thus we can apply induction to show the summation is nonnegative.

**Case 2:** Assume that $j + 1 \in S$ and thus $a = \min S = j + 1$. For this case we directly analyze the term of interest. First break up the product $E_S$ so that

$$\sum_{\pi \in \Sigma_N} W(\pi)\left(m_j - x_{\pi(j)}\right) E_S(\pi) G(\pi) = \sum_{\pi \in \Sigma_N} W(\pi)\left(m_j - x_{\pi(j)}\right) E_{\{j+1\}}(\pi) E_{S \setminus \{j+1\}}(\pi) G(\pi).$$

Then define $W_j(\pi) \equiv \frac{x_{\pi(j)}}{m_{j+1}(\pi)} \geqslant 0$. Notice that $W_j$ is $j$-invariant. Moreover, observe that $\widetilde{W}(\pi) \equiv W(\pi) \times W_j(\pi)$ is $j$-invariant. Then by Lemma 4.7, we have

$$\left(m_j(\pi) - x_{\pi(j)}\right) E_{\{j+1\}}(\pi) = \frac{r_j - 1}{r_j} \left\{ \left(x_{\pi(j+1)} - x_{\pi(j)}\right) + W_j(\pi)\left(m_{j+1}(\pi) - x_{\pi(j+1)}\right) \right\}.$$

Applying the identity to the summation and ignoring the nonnegative constant $(r_j - 1)/r_j$ gives

$$\begin{aligned}&\sum_{\pi\in\Sigma_N} W(\pi)\left((x_{\pi(j+1)} - x_{\pi(j)}) + W_j(\pi)\left(m_{j+1}(\pi) - x_{\pi(j+1)}\right)\right) E_{S\setminus\{j+1\}}G(\pi)\\ &= \underbrace{\sum_{\pi\in\Sigma_N} W(\pi)\left(x_{\pi(j+1)} - x_{\pi(j)}\right) E_{S\setminus\{j+1\}}(\pi)G(\pi)}_{\text{III}} + \underbrace{\sum_{\pi\in\Sigma_N} \widetilde{W}(\pi)\left(m_{j+1}(\pi) - x_{\pi(j+1)}\right) E_{S\setminus\{j+1\}}G(\pi)}_{\text{IV}}\end{aligned}$$

For term III, we can set $F(\pi) = W(\pi)E_{S\setminus\{j+1\}}(\pi)$ as for this choice of $F$, $F(s_j \circ \pi) = F(\pi)$. Then apply Lemma 4.1 to show that the summation is nonnegative. Induction demonstrates that IV is nonnegative. $\square$

We now apply Lemma 4.2 to the symmetric polynomials. Denote $B_k^{(n)} := S_k^{(n)}/\binom{n}{k}$ where $S_k^{(j)} = S_k(E_1, \dots, E_j)$

**Lemma 4.3.** *$(B_k)_{k=0}^n$ is a nonnegative demi-supermartingale, meaning that for every nonnegative coordinatewise decreasing $G$,*

$$\mathbb{E}[B_k] = 1 \quad \textit{and} \quad \mathbb{E}[(B_{k+1} - B_k)\, G(B_1, \dots, B_k)] \leqslant 0.$$

*Proof of Lemma 4.3.* We have by Lemma 4.8 that

$$\binom{j}{k}(j-k)\left(B_{k+1}^{(j)} - B_k^{(j)}\right) = (k+1)S_{k+1}^{(j)} - (j-k)S_k^{(j)}.$$

Taking $j = n$ gives the expression of interest

$$\begin{aligned}(k+1)S_{k+1} - (n-k)S_k &= \sum_{|S|=k}\sum_{i\notin S} E_i E_S - \sum_{|S|=k}\sum_{i\notin S} E_S\\ &= \sum_{|S|=k}\sum_{i\notin S} E_S(E_i - 1).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}[(B_{k+1} - B_k)\, G] &= \binom{n}{k}^{-1}(n-k)^{-1}(N!)^{-1}\sum_{\pi\in\Sigma_N}\sum_{|S|=k}\sum_{i\notin S} E_S(x_{\pi(i)} - m_i)(1/m_i)G(\pi)\\ &= \binom{n}{k}^{-1}(n-k)^{-1}(N!)^{-1}\sum_{|S|=k}\sum_{i\notin S}\sum_{\pi\in\Sigma_N} E_S(x_{\pi(i)} - m_i)(1/m_i)G(\pi)\\ &= \binom{n}{k}^{-1}(n-k)^{-1}(N!)^{-1}\sum_{|S|=k}\sum_{i\notin S}\sum_{\pi\in\Sigma_N} E_{S\cap\{i+1,\dots,n\}}E_{S\cap[i-1]}(x_{\pi(i)} - m_i)(1/m_i)G(\pi).\end{aligned}$$

Clearly $S \cap \{i+1, \dots, n\} \subset \{i+1, \dots, N\}$. Moreover, $E_{S\cap[i-1]}$ and $1/m_i$ are both $(i-1)$-invariant and nonnegative. Define

$$W(\pi) = E_{S\cap[i-1]}(\pi)/m_i(\pi).$$

Finally, $G(\pi)$ is swap-decreasing by the composition lemma in Lemma 4.4. Applying Lemma 4.2 gives the upper bound. $\square$

With the result that the SymPol-WoR process is a demi-supermartingale, we can prove Theorem 3.1 by noting that $1\{\max_{0\leqslant k\leqslant n} B_k(\mathbf{X}) \geqslant 1/\alpha\}$ is coordinate increasing.

### 4.1 Auxiliary lemmas

**Lemma 4.4.** *Let $G$ be coordinatewise decreasing and for each $f_i$ in $(f_1, \ldots, f_k)$ assume $f_i$ is swap-increasing. Then $G(f_1, \ldots, f_k)$ is swap-decreasing.*

*Proof of Lemma 4.4.* Let $\pi \preceq \pi'$. Then $f_i(\pi') \geqslant f_i(\pi)$ for every $i$. Therefore,

$$G\left(f_1(\pi'), \ldots, f_k(\pi')\right) \leqslant G\left(f_1(\pi), \ldots, f_k(\pi)\right),$$

so $G(f_1, \ldots, f_k)$ is swap-decreasing. $\square$

**Lemma 4.5.** *$S_k(\pi)$ is swap-increasing and hence $B_k$ is swap-increasing.*

*Proof of Lemma 4.5.* It is sufficient to show that if $x_{\pi(j)} \leqslant x_{\pi(j+1)}$ then

$$S_k(\pi) \leqslant S_k(s_j \circ \pi).$$

The first step is to prove an identity. Let $S_k^- \equiv S_k(E_1, \ldots, E_{j-1}, E_{j+2}, \ldots, E_n)$, meaning it is the degree-$k$ symmetric polynomial after removing the $E_j, E_{j+1}$ variables. Then by Lemma 4.10

$$S_k = E_j E_{j+1} S_{k-2}^- + (E_j + E_{j+1}) S_{k-1}^- + S_k^- \tag{13}$$

With the identity in (13), it is sufficient to look only at the product and sum of the two e-variables of interest $E_j E_{j+1}$ and $E_j + E_{j+1}$.

We compute the product first.

$$E_j E_{j+1} = \frac{x_{\pi(j)}}{m_j(\pi)} \frac{x_{\pi(j+1)}}{m_{j+1}(\pi)} = x_{\pi(j)} x_{\pi(j+1)} \frac{r_j r_{j+1}}{T_j(\pi) T_{j+1}(\pi)}. \tag{14}$$

Multiplication is commutative and $T_j(\pi)$ is $j$-invariant. Thus the only term that changes is $T_{j+1}(\pi) = T_j - x_{\pi(j)}$. If $s_j$ is an up-move then $x_{s_j \circ \pi(j)}$ increases and $T_{j+1}(s_j \circ \pi)$ decreases meaning that the product $E_j E_{j+1}$ increases.

Next we compute the summation, which is more involved. We first write the identity

$$E_j + E_{j+1} = \underbrace{(r-1)\left(1 - \frac{T_{j+2}}{T_j}\right)}_{\mathrm{I}} + \frac{1}{r} \underbrace{(E_j + E_j E_{j+1})}_{\mathrm{II}}.$$

This identity can be shown using $(1 - E_j/r)(1 - E_{j+1}/(r-1)) = T_{j+2}/T_j$. Then the first term I is $s_j$-invariant. The second term II is swap-increasing since we showed that the product is swap-increasing and $E_j$ increases on an up-move.

Notice that $T_j(\pi)$ appears in the denominator in (14) and thus we should require $T_j(\pi) > 0$. Indeed, we do and otherwise set $E_j(\pi) = 1$ by convention. $\square$

**Lemma 4.6.** *For any $1 \leqslant a \leqslant N-1$,*

$$r_a(m_a - X_a) = \sum_{i=a}^{N-1} (N-i)(X_{i+1} - X_i).$$

*Taking $a = 1$ gives the full decomposition*

$$N(m - X_1) = \sum_{i=1}^{N-1} (N-i)(X_{i+1} - X_i).$$

*Proof of Lemma 4.6.*

$$
\begin{aligned}
r_a(m_a - X_a) &= r_a m_a - r_a X_a \\
&= T_a - (N - a + 1)X_a \\
&= \sum_{i=a+1}^{N} (X_i - X_a) \\
&= \sum_{i=a+1}^{N} \sum_{j=a}^{i-1} (X_{j+1} - X_j) \\
&= \sum_{j=a}^{N-1} (X_{j+1} - X_j)(N - j),
\end{aligned}
$$

where the last step interchanges order of summation. □

**Lemma 4.7.** *For $W_j(\pi) \equiv x_{\pi(j)}/m_{j+1}(\pi)$ we have*

$$
\left(m_j(\pi) - x_{\pi(j)}\right) E_{j+1}(\pi) = \frac{r_j - 1}{r_j} \left\{ \left(x_{\pi(j+1)} - x_{\pi(j)}\right) + W_j(\pi) \left(m_{j+1}(\pi) - x_{\pi(j+1)}\right) \right\}
$$

*Proof of Lemma 4.7.*

$$
\begin{aligned}
r_j m_j(\pi) = T_j &= T_{j+1} + x_{\pi(j)} \\
&= (r_j - 1)\, m_{j+1}(\pi) + x_{\pi(j)}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
m_j(\pi) - x_{\pi(j)} &= \frac{(r_j - 1)m_{j+1}(\pi) + x_{\pi(j)} - r_j x_{\pi(j)}}{r_j} \\
&= \frac{r_j - 1}{r_j} \left(m_{j+1}(\pi) - x_{\pi(j)}\right)
\end{aligned}
$$

Focusing on $m_{j+1}(\pi) - x_{\pi(j)}$ and multiplying by $E_{j+1}$ gives

$$
\begin{aligned}
\left(m_{j+1}(\pi) - x_{\pi(j)}\right) E_{j+1} &= \left(x_{\pi(j+1)} - x_{\pi(j)} E_{j+1}\right) \\
&= \left(x_{\pi(j+1)} - x_{\pi(j)}\right) + x_{\pi(j)} \left(1 - E_{j+1}\right) \\
&= \left(x_{\pi(j+1)} - x_{\pi(j)}\right) + W_j(\pi) \left(m_{j+1}(\pi) - x_{\pi(j+1)}\right),
\end{aligned}
$$

for $W_j(\pi) \equiv x_{\pi(j)}/m_{j+1}(\pi)$. □

**Lemma 4.8.** *Let $B_k^{(j)} = S_k^{(j)}/\binom{j}{k}$ where $S_k^{(j)} = S_k(E_1, \ldots, E_j)$. Then*

$$
\binom{j}{k}(j-k)\left(B_{k+1}^{(j)} - B_k^{(j)}\right) = (k+1)S_{k+1}^{(j)} - (j-k)S_k^{(j)}.
$$

*Proof of Lemma 4.8.* By definition $S_k^{(j)} = B_k^{(j)}\binom{j}{k}$. Therefore

$$
\begin{aligned}
\binom{j}{k}\left(B_{k+1}^{(j)} - B_k^{(j)}\right) &= \left(\frac{\binom{j}{k}}{\binom{j}{k+1}} S_{k+1}^{(j)} - S_k^{(j)}\right) \\
&= \frac{k+1}{j-k} S_{k+1}^{(j)} - S_k^{(j)},
\end{aligned}
$$

Multiplying by $(j-k)$ gives

$$\binom{j}{k}(j-k)\left(B_{k+1}^{(j)}-B_k^{(j)}\right)=(k+1)S_{k+1}^{(j)}-(j-k)S_k^{(j)}.$$

□

**Lemma 4.9.** *We state as a fact the recursive definition of the elementary symmetric polynomials. We have*

$$S_k(E_1,\dots,E_n)=S_k(E_1,\dots,E_{n-1})+S_{k-1}(E_1,\dots,E_{n-1})E_n$$

**Lemma 4.10.** *Let* $S_k^{-\{i,i+1\}}\equiv S_k(E_1,\dots,E_{i-1},E_{i+2},\dots,E_n)$. *Then*

$$S_k=E_iE_{i+1}S_{k-2}^{-\{i,i+1\}}+(E_i+E_{i+1})\,S_{k-1}^{-\{i,i+1\}}+S_k^{-\{i,i+1\}}$$

*Proof.* Using the recursive definition in Lemma 4.9 we can peel off $E_i$ to obtain

$$S_k(E_1,\dots,E_n)=S_k(E_1,\dots,E_{i-1},E_{i+1},\dots,E_n)+S_{k-1}(E_1,\dots,E_{i-1},E_{i+1},\dots,E_n)E_i \tag{15}$$

Peeling off $E_{i+1}$ from each term in (15) yields the result. □

# 5 Proof of SymPol's validity

In this section, we prove that applying SymPol to the naïve e-variables is still valid. Write $\widetilde{E}_i=X_i/m$, where $m=N^{-1}\sum_{i=1}^N x_i$, and define

$$B_k\equiv B_k(\widetilde{E}_1,\dots,\widetilde{E}_n)$$

after sampling $X_1,\dots,X_n$ without replacement. It suffices to prove that $(B_k)_{k=0}^n$ forms a demi-supermartingale; that is, for every nonnegative coordinatewise decreasing $G$,

$$\mathbb{E}[(B_{k+1}-B_k)G(B_1,\dots,B_k)]\leqslant 0.$$

In this section, the data are exchangeable and the only randomness comes from the random subsets rather than from random orderings, as in the proof in Section 4. This means we can think of both $B_k$ and $G$ as functions from subsets $S$ to nonnegative numbers.

We will preprocess the data in two ways. We will first relabel the population so that it is in increasing order:

$$x_1\leqslant x_2\leqslant\cdots\leqslant x_N.$$

In addition, we will normalize the data so that the mean is 1. This implies

$$\sum_{i=1}^N(x_i-1)=0.$$

**Step 1: Calculating $\mathbb{E}[B_k]$ for a fixed $k$** Begin by fixing $k$ and some nonnegative $G$. We begin by computing $\mathbb{E}[B_k G]$.

**Lemma 5.1.** *Let $G \geqslant 0$ and $x_S = \prod_{i \in S} x_i$. Then*

$$\binom{n}{k}\binom{N}{n}\mathbb{E}[B_k G] = \sum_{|S|=k} x_S w_S \quad \text{where} \quad w_S = \sum_{|\mathcal{J}|=n,\, \mathcal{J} \supset S} G(\mathcal{J}).$$

*Proof of Lemma 5.1.* The proof simply changes the order of summation.

$$\begin{aligned}
\mathbb{E}[B_k G] &= \binom{N}{n}^{-1} \sum_{|\mathcal{J}|=n} B_k(\mathcal{J}) G(\mathcal{J}) \\
&= \binom{N}{n}^{-1} \sum_{|\mathcal{J}|=n} G(\mathcal{J}) \binom{n}{k}^{-1} \sum_{|S|=k,\, \mathcal{J} \supset S} x_S \\
&= \binom{N}{n}^{-1} \binom{n}{k}^{-1} \sum_{|S|=k} x_S \underbrace{\sum_{|\mathcal{J}|=n,\, \mathcal{J} \supset S} G(\mathcal{J})}_{:= w_S} \\
&= \binom{N}{n}^{-1} \binom{n}{k}^{-1} \sum_{|S|=k} x_S w_S.
\end{aligned}$$

□

**Step 2: Calculating the Difference in Expectations** Next, we compute the expectation of the difference in symmetric polynomials $(B_{k+1} - B_k)G$.

**Lemma 5.2.** *We have*

$$\mathbb{E}[(B_{k+1} - B_k)G] = \frac{1}{(k+1)\binom{n}{k+1}\binom{N}{n}} \sum_{|S|=k+1} w_S \beta_S,$$

*where*

$$\beta_S = (k+1)x_S - \sum_{i \in S} x_{S \setminus i}$$

*Proof.* We begin by noting for a fixed subset $S$, where $|S| = k$, we have

$$\begin{aligned}
\sum_{i \notin S} w_{S \cup i} &= \sum_{i \notin S} \sum_{\mathcal{J} : |\mathcal{J}|=n, \mathcal{J} \supset S \cup i} G(\mathcal{J}) \\
&= \sum_{i \notin S, i \in \mathcal{J}} \sum_{\mathcal{J} : |\mathcal{J}|=n, \mathcal{J} \supset S} G(\mathcal{J}) \\
&= (n-k) w_S.
\end{aligned}$$

The math above shows that two sets are equivalent. The first set is constructed if we first fix a set $S$ and index $i \notin S$ then sum over all $\mathcal{J}$ containing both $i$ and $S$. The second set is constructed as we first fix a set $S$ and $\mathcal{J}$ and then sum over all $i \notin S$ and $i \in \mathcal{J}$. Lastly, we interchange sums and note $|\mathcal{J} \setminus S| = n - k$.

We continue with some more algebra:

$$\begin{aligned}\sum_{|S|=k} x_S w_S &= \frac{1}{n-k}\sum_{|S|=k} x_S \sum_{i\notin S} w_{S\cup i}\\ &= \frac{1}{n-k}\sum_{|\widetilde{S}|=k+1} w_{\widetilde{S}} \sum_{i\in\widetilde{S}} x_{\widetilde{S}\setminus i},\end{aligned}$$

where the second identity changes the order of summation.

Then directly compute the difference,

$$\begin{aligned}\binom{N}{n}\left(\mathbb{E}[B_{k+1}G]-\mathbb{E}[B_kG]\right) &= \binom{n}{k+1}^{-1}\sum_{|\widetilde{S}|=k+1} x_{\widetilde{S}}w_{\widetilde{S}} - \binom{n}{k}^{-1}\sum_{|S|=k} x_S w_S\\ &= \binom{n}{k+1}^{-1}\sum_{|\widetilde{S}|=k+1} x_{\widetilde{S}}w_{\widetilde{S}} - \binom{n}{k}^{-1}\frac{1}{n-k}\sum_{|\widetilde{S}|=k+1} w_{\widetilde{S}}\sum_{i\in S_1} x_{\widetilde{S}\setminus i}\\ &= \sum_{|\widetilde{S}|=k+1} w_{\widetilde{S}}\left(x_{\widetilde{S}}\binom{n}{k+1}^{-1} - \binom{n}{k}^{-1}\frac{1}{n-k}\sum_{i\in\widetilde{S}} x_{\widetilde{S}\setminus i}\right).\end{aligned}$$

Lastly, multiply both sides by $(k+1)\binom{n}{k+1} = (n-k)\binom{n}{k}$ to obtain

$$(k+1)\binom{n}{k+1}\binom{N}{n}\mathbb{E}[(B_{k+1}-B_k)\,G] = \sum_{|\widetilde{S}|=k+1} w_{\widetilde{S}}\left((k+1)x_{\widetilde{S}} - \sum_{v\notin\widetilde{S}} x_{\widetilde{S}\setminus i}\right) \tag{16}$$

□

**Step 3: Upper bounding the weighted sum** The last step is to show that (16) $\leqslant 0$. This shows that $(B_k)_{k=0}^n$ is a demi-supermartingale and thus that naïve SymPol is valid under a without-replacement model.

**Lemma 5.3.**

$$\sum_{|S|=k+1} w_S\left((k+1)x_S - \sum_{v\in S} x_{S\setminus v}\right) \leqslant 0.$$

*Proof of Lemma 5.3.* We first exchange the order of summation to write

$$\begin{aligned}\sum_{|S|=k+1} w_S\left((k+1)x_S - \sum_{i\in S} x_{S\setminus i}\right) &= \sum_{j=1}^{N}(x_j-1)\sum_{|S|=k,j\in S} w_S x_{S\setminus j}\\ &= \sum_{i=1}^{N}(x_j-1)\,\Lambda_j \end{aligned}\tag{17}$$

where we define $\sum_{|S|=k+1,j\in S} w_S x_{S\setminus j} = \Lambda_j$. Next, we apply Chebyshev's summation inequality. Chebyshev's sum inequality also appears in the proof of [6], and thus seems central to SymPol proofs.

By Lemma 5.4 $\Lambda_i$ are nonincreasing in $i$. Moreover, the population is nondecreasing by construction ($x_1 \leqslant \cdots \leqslant x_N$). Thus

$$\begin{aligned} N^{-1} \sum_{i=1}^{N} x_i \Lambda_i &\leqslant \left( N^{-1} \sum_{i=1}^{N} x_i \right) \left( N^{-1} \sum_{i=1}^{N} \Lambda_i \right) \\ &= N^{-1} \sum_{i=1}^{N} \Lambda_i. \end{aligned}$$

This implies that (17) $\leqslant 0$ and thus that (16) $\leqslant 0$. □

**Lemma 5.4.** *Let $\Lambda_i$ be defined as*

$$\Lambda_i := \sum_{|S|=k+1, i \in S} \sum_{|\mathcal{J}|=n, S \subset \mathcal{J}} G(\mathcal{J}) \prod_{j \in S \setminus i} x_j.$$

*Then*

$$\Lambda_1 \geqslant \Lambda_2 \geqslant \cdots \geqslant \Lambda_N.$$

*Proof of Lemma 5.4.* Fix $i < j$, so $x_i \leqslant x_j$ by construction of the population. Then exchange the order of summation and separate the terms according to whether $j \in \mathcal{J}$ to obtain

$$\begin{aligned} \Lambda_i - \Lambda_j = &\sum_{\substack{|\mathcal{J}|=n-1 \\ i,j \notin \mathcal{J}}} [G(\mathcal{J} \cup \{i\}) - G(\mathcal{J} \cup \{j\})] \sum_{\substack{|S|=k \\ S \subset \mathcal{J}}} x_S \\ &+ \sum_{\substack{|\mathcal{J}|=n-2 \\ i,j \notin \mathcal{J}}} G(\mathcal{J} \cup \{i,j\})(x_j - x_i) \sum_{\substack{|S|=k-1 \\ S \subset \mathcal{J}}} x_S. \end{aligned}$$

The first sum is nonnegative because replacing $x_i$ by $x_j$ increases each of $B_1, \ldots, B_k$, while $G$ is coordinatewise nonincreasing. The second is nonnegative because $G \geqslant 0$ and $x_i \leqslant x_j$. Hence $\Lambda_i \geqslant \Lambda_j$ for every $i < j$. □

# 6 Discussion

When the data are sampled without replacement, the naïve SymPol interval remains valid while the Gaffke interval fails. Moreover, one can use the structure of sampling without replacement to obtain SymPol-WoR intervals that are tight across a variety of competing bounds. These intervals satisfy many desired properties including adaptivity to the variance, asymmetry, and ability to shrink as $n \to N$. We hope this inspires more work on confidence intervals under sampling without replacement and on exploring SymPol and Gaffke's interval for nonasymptotic inference.

**Acknowledgments** The author thanks Ian Waudby-Smith for helpful discussions. Claude Opus 5 gave the proofs in Section 4 and Section 5 and GPT 5.6 gave the counterexample to Gaffke's interval in Theorem 3.2. The author verified and modified the AI proofs and takes full responsibility for the content.

# References


[1] Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015. 2, 4, 5, 7

[2] N. Gaffke. Three test statistics for a nonparametric one-sided hypothesis on the mean of a nonnegative variable. *Mathematical Methods of Statistics*, 14(4):451–467, 2005. 1

[3] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963. 2, 4

[4] Xinran Li and Peng Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520): 1759–1769, 2017. 3

[5] Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990. 2

[6] Jiahao Ming, Aaditya Ramdas, Yi Shen, Ruodu Wang, and Ian Waudby-Smith. Combining e-values using demi-supermartingales, 2026. 1, 2, 6, 16

[7] Jiahao Ming, Aaditya Ramdas, Yi Shen, Ruodu Wang, and Ian Waudby-Smith. Gaffke's confidence interval for the mean of bounded data is inadmissible but asymptotically efficient. *arXiv preprint arXiv:2607.18661*, 2026. 2, 8

[8] B. L. S. Prakasa Rao. *Associated Sequences, Demimartingales and Nonparametric Inference*. Probability and Its Applications. Birkhäuser, Basel, 2012. 2

[9] Aaditya Ramdas and Ruodu Wang. Hypothesis testing with e-values. *Foundations and Trends® in Statistics*, 1(1-2):1–390, 2025. 1

[10] Robert J Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, pages 39–48, 1974. 2, 4, 5

[11] Roman Vershynin. High-dimensional probability. *Cambridge Series in Statistical and Probabilistic Mathematics*, 47, 2019. 4

[12] Nikos Vlassis and Philip S Thomas. An Exact Distribution-Free Test for Means of Nonnegative Random Variables. *arXiv preprint arXiv:2607.08415*, 2026. 1, 2

[13] Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736–1754, 2021. 1

[14] Weizhen Wang and Linda H Zhao. Nonparametric tests for the mean of a non-negative population. *Journal of statistical planning and inference*, 110(1-2):75–96, 2003. 1

[15] Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):1–27, 2024. 4, 5, 6, 7