Effective Sample Size for Importance Sampling based on the discrepancy measures

L. Martino[⋆], V. Elvira[◇], F. Louzada[⋆]
[⋆] Universidade de São Paulo, São Carlos (Brazil).
[◇] Universidad Carlos III de Madrid, Leganés (Spain).

Abstract

The Effective Sample Size (ESS) is an important measure of efficiency of Monte Carlo methods such as Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) techniques. In IS context, an approximation $\widehat{ESS}$ of the theoretical ESS definition is widely applied, involving the sum of the squares of the normalized importance weights. This formula $\widehat{ESS}$ has become an essential piece within Sequential Monte Carlo (SMC) methods using adaptive resampling procedures. The expression $\widehat{ESS}$ is related to the Euclidean distance between the probability mass described by the normalized weights and the discrete uniform probability mass function (pmf). In this work, we derive other possible ESS functions based on different discrepancy measures between these two pmfs. Several examples are provided involving, for instance, the geometric and harmonic means of the weights, the discrete entropy (including the *perplexity* measure, already proposed in literature) and the Gini coefficient. We list five requirements which a generic ESS function should satisfy, allowing us to classify different ESS measures. We also compare the most promising ones by means of numerical simulations.

1 Introduction

Sequential Monte Carlo (SMC) methods (a.k.a., particle filtering algorithms) are important tools for Bayesian inference [8], extensively applied in signal processing [7, 15, 21] and statistics [9, 23, 24]. A key point for the success of a SMC method is the use of resampling procedures, applied for avoiding the degeneracy of the importance weights [7, 9]. However, the application of resampling increases the variance of the Monte Carlo estimators so that one desire to employ resampling steps parsimoniously, only when it is strictly required. This adaptive implementation of the resampling procedure needs the use of the concept of Effective Sample Size (ESS) [7, 19, 24].

ESS is an important concept for measuring the efficiency of different Monte Carlo methods, such as Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) techniques [4, 13, 19, 24]. ESS is theoretically defined as the equivalent number of independent samples generated directly form the target distribution, which yields the same efficiency in the estimation obtained by the MCMC or IS algorithms. Thus, a possible mathematical definition [13, 17] considers the ESS function proportional to the ratio between the variance of the ideal Monte Carlo estimator (drawing samples directly from the target) over the variance of the estimator obtained by MCMC or IS techniques, using with the same number of samples in both estimators.

The most common choice in literature to approximate this theoretical ESS definition is the formula $\widehat{ESS} \approx \frac{1}{\sum_{n=1}^{M} \bar{w}_n^2}$, which involves (only) the normalized importance weights $\bar{w}_n$, $n = 1, \ldots, N$ [7, 9, 18, 24]. This expression, obtained after several approximations of the definition, presents different weaknesses discussed in this work. Another measure called *perplexity*, involving the discrete entropy [5] of the weights has been also proposed in [1]; see also [24, Chapter 4], [10, Section 3.5].

However, the ESS approximation $\widehat{ESS}$ is widely used in practice, since it is easily to be applied and provides good performance, in general. It is possible to show that $\widehat{ESS}$ is related to discrepancy between the multinomial probability mass function (pmf) defined by the normalized weights $\bar{w}_n$, $n = 1, \ldots, N$, and the discrete uniform pmf. When the pmf defined by $\bar{w}_n$ is close to the discrete uniform pmf, $\widehat{ESS}$ provides high values otherwise, when the pmf defined by $\bar{w}_n$ is concentrated mainly in one weight, $\widehat{ESS}$ provides small values. More specifically, we show that $\widehat{ESS}$ is related to the Euclidean distance between pmfs. Moreover, we deduce other possible ESS functions based on different distances. Then, we describe and discuss five requirements, three strictly needed and two welcome conditions, that a generalized ESS (G-ESS) function should satisfy. Several examples, involving for instance geometric and harmonic means, discrete entropy [5] and the Gini coefficient [14, 20], are presented. Additionally, four families of proper G-ESS functions are designed. We classify the novel G-ESS functions (including also the *perplexity* measure [1, 24]) according to the conditions fulfilled. We focus our attention in the G-ESS functions which satisfies all the desirable conditions and compare them by means of numerical simulations. This analysis shows that at least one novel G-ESS expression, defined as the inverse of the maximum of the normalized weights, $\frac{1}{\max[\bar{w}_1, \ldots, \bar{w}_N]}$, presents interesting features from a theoretical and practical point of view and it can be considered a valid alternative to the standard formula $\frac{1}{\sum_{n=1}^{M} \bar{w}_n^2}$.

2 Effective Sample Size for Importance Sampling

Let us denote the target probability density function (pdf) as $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$ (known up to a normalizing constant) with $\mathbf{x} \in \mathcal{X}$. Moreover, we consider the following integral involving $\bar{\pi}(\mathbf{x})$ and a square-integrable function $h(\mathbf{x})$,

$$I = \int_{\mathcal{X}} h(\mathbf{x}) \bar{\pi}(\mathbf{x}) d\mathbf{x}, \tag{1}$$

which we desire to approximate using a Monte Carlo approach. If we are able to draw N independent samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from $\bar{\pi}(\mathbf{x})$, then the Monte Carlo estimator of I is

$$\widehat{I} = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{x}_n) \approx I, \tag{2}$$

where $\mathbf{x}_n \sim \bar{\pi}(\mathbf{x})$, with $n = 1, \ldots, N$. However, in general, generating samples directly from the target, $\bar{\pi}(\mathbf{x})$, is impossible. Alternatively, we can draw N samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from a (simpler) proposal pdf $q(\mathbf{x})$,[1] and then assign a weight to each sample, $w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$, with $n = 1, \ldots, N$, according to the importance sampling (IS) approach. Defining the normalized weights,

$$\bar{w}_n = \frac{w_n}{\sum_{i=1}^N w_i}, \quad n = 1, \ldots, N, \tag{3}$$

then the IS estimator is

$$\widetilde{I} = \sum_{n=1}^N \bar{w}_n h(\mathbf{x}_n) \approx I, \tag{4}$$

with $\mathbf{x}_n \sim q(\mathbf{x})$, $n = 1, \ldots, N$. In general, the estimator $\widetilde{I}$ is less efficient than $\widehat{I}$, since the samples are not directly generate by $\bar{\pi}(\mathbf{x})$. In several applications [7, 9, 15, 21], it is necessary to measure in some way the efficiency that we lose using $\widetilde{I}$ instead of $\widehat{I}$. The idea is to define the Effective Sample Size (ESS) as ratio of the variances of the estimators [17],

$$ESS = N \frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}. \tag{5}$$

Finding a useful expression of ESS derived analytically from the theoretical definition above is not straightforward. Then, different derivations [17, 18], [9, Chapter 11], [24, Chapter 4] proceed using several approximations and assumptions for yielding an expression useful from a practical point of view. A well-known formula, widely used in literature [9, 19, 24], is

$$\widehat{ESS} = P_N^{(2)}(\bar{\mathbf{w}}), \tag{6}$$

$$= \frac{1}{\sum_{i=1}^N \bar{w}_n^2} = \frac{\left(\sum_{i=1}^N w_n\right)^2}{\sum_{i=1}^N w_n^2}, \tag{7}$$

where we have used the the normalized weights $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$ in the first equality, and the unnormalized one in the second equality. The reason of using the notation $P_N^{(2)}(\bar{\mathbf{w}})$ will appear clear later (the subindex N denotes the number of weights involved, and the reason of the super-index will be clarified in Section 4.3). An interesting property of the expression (7) is that

$$1 \leq P_N^{(2)}(\bar{\mathbf{w}}) \leq N. \tag{8}$$

Below, we discuss some limitations of the formula $P_N^{(2)}(\bar{\mathbf{w}})$.

2.1 Discussion about $P_N^{(2)}$

Although, the formula $P_N^{(2)}(\bar{\mathbf{w}}) = \frac{1}{\sum_{i=1}^N \bar{w}_n^2}$ is widely used and performs reasonable well in different scenarios, the formula $P_N^{(2)}$ presents several weaknesses. First of all, the starting definition of

[1] We assume that $q(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{X}$, and $q(\mathbf{x})$ has heavier tails than $\bar{\pi}(\mathbf{x})$.

effective sample size, $ESS = N\frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}$, does not take in account the bias due to the use of self-normalized weights. Thus, a more complete definition could be

$$ESS = N\frac{\text{MSE}\left[\widehat{I}\right]}{\text{MSE}\left[\widetilde{I}\right]} = N\frac{\text{var}_\pi\left[\widehat{I}\right]}{\text{MSE}\left[\widetilde{I}\right]}, \tag{9}$$

where we have considered the Mean Square Error (MSE) and we have taken into account that $\widehat{I}$ is unbiased. Furthermore, several approximations have been applied to arrived to the final expression (for instance, the author in [17] applies twice the delta method [3]). Owing to all the employed approximations, the final formula $P_N^{(2)}$ does not depend on the particles $\mathbf{x}_n$, $n = 1, \dots, N$, which is obviously a drawback since we are trying to measure the effective sample size of the set of weighted particles. Figure 1 shows the progressive loss of information that we have first normalizing the weights and then removing the information related to the position of the particles. Conversely to the previous observation, the fact the $P_N^{(2)}$ is independent from the function $h(\mathbf{x})$ is positive and desirable (starting from the definition $ESS = N\frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}$, this is not a trivial conclusion). Another important issue is that all the derivation of $P_N^{(2)}$ has been development under the assumption of a single proposal context, i.e.,

$$\mathbf{x}_n \sim q(\mathbf{x}), \quad n = 1, \dots, N,$$

but in general the formula $P_N^{(2)}$ is applied in a multiple proposal setting, i.e.,

$$\mathbf{x}_1 \sim q_1(\mathbf{x}), \mathbf{x}_2 \sim q_2(\mathbf{x}), \dots, \mathbf{x}_N \sim q_N(\mathbf{x}),$$

which is the case of Population Monte Carlo (PMC) [2], particle filtering [7] and sequential Monte Carlo [9] methods, more generally. Finally, note that $1 \leq P_N^{(2)} \leq N$ which can appear an interesting feature after a first examination, but actually it does not encompass completely the theoretical consequences included in the general definition $ESS = N\frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}$. Indeed, by this general definition of ESS, we have

$$0 \leq ESS \leq B, \qquad B \geq N,$$

i.e., namely ESS can be less than 1, when $\text{var}_q[\widetilde{I}] >> \text{var}_\pi[\widehat{I}]$, and even greater than N, when $\text{var}_q[\widetilde{I}] < \text{var}_\pi[\widehat{I}]$: this case occurs when *negative correlation* is induced among the generated samples [11].

However, $P_N^{(2)}$ has become an essential piece within Monte Carlo methods using adaptive resampling steps [7, 15, 21]. It has been widely accepted as a valid measure of effective sample size. Below, we describe a simple justification for the diffusion of $P_N^{(2)}$ as a suitable ESS approximation in an adaptive resampling context.

3 Discrepancy measures for adaptive resampling

Many population Monte Carlo (PMC) [2, 12] or sequential Monte Carlo (SMC) methods [9, 15], employ resampling steps for updating the parameters of the used proposal functions. On the one

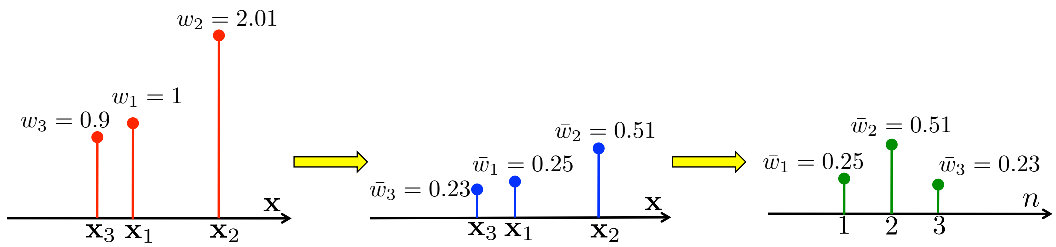


Figure 1: Graphical representation of the loss of statistical information normalizing the weights and removing the positions of the particles ($N = 3$).

hand, PMC and SMC suffer the so-called path degeneracy, i.e., after some iterations only one sample is statistically relevant in terms of importance weights. This problem could be solved applying resampling procedures. However, on the other hand, the application of resampling yields loss of diversity in the set of samples, incorporating additional variance in the Monte Carlo estimators. Therefore, one often attempts to apply resampling steps only in certain specific iterations, when it is considered strictly required.

In a standard multinomial resampling, the indices of the particles used in the next generation are drawn according to a multinomial probability mass function (pmf) defined by the normalized weights $\bar{w}_n = \frac{w_n}{\sum_{i=1}^N w_i}$, with $n = 1, \ldots, N$. Ideally, if the samples are drawn directly from the target distribution all the weights w_n are equal, so that $\bar{w}_n = \frac{1}{N}$, $n = 1, \ldots, N$. We denote

$$\bar{\mathbf{w}}^* = \left[\frac{1}{N}, \ldots, \frac{1}{N}\right], \tag{10}$$

that is the vector with equal components $\bar{w}_n = \frac{1}{N}$, $n = 1, \ldots, N$. It important to note that the inverse is not always true: namely the scenario $\bar{w}_n = \frac{1}{N}$, $n = 1, \ldots, N$, could occur even if the proposal density is different from the target. Hence, in general, in this case we can assert $ESS \leq N$ (considering independent, non-negative correlated, samples). The other extreme case is

$$\bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_N = 0], \tag{11}$$

i.e., $\bar{w}_j = 1$ and $\bar{w}_n = 0$ (it can occurs only if $\pi(\mathbf{x}_n) = 0$), for $n \neq j$ with $j \in \{1, \ldots, N\}$. In general, in this case $ESS \leq 1$.

Optimistic approach. Using only the information provided by the vector $\bar{\mathbf{w}}$, one can try to deduce when the application of resampling is required using the previous considerations. In this context, the function $P_N^{(2)}(\bar{\mathbf{w}})$ appears useful. More specifically, the function $P_N^{(2)}(\bar{\mathbf{w}})$ employ an *optimistic approach* for the two extreme cases previously above:

$$P_N^{(2)}(\bar{\mathbf{w}}^*) = N, \tag{12}$$

$$P_N^{(2)}(\bar{\mathbf{w}}^{(j)}) = 1, \quad \forall j \in \{1, \ldots, N\}. \tag{13}$$

Moreover, considering a vector of type

$$\bar{\mathbf{w}} = \left[0, \frac{1}{C}, \frac{1}{C}, 0, \dots, 0, 0, \frac{1}{C}, \dots, 0\right],$$

where only C entries are non-null with the same weight $\frac{1}{C}$, note that

$$P_N^{(2)}(\bar{\mathbf{w}}) = C. \tag{14}$$

Figure 2 summarizes graphically these cases. The underlying idea behind the use of $P_N^{(2)}$ is that if the pmf $\bar{w}_n$, $n = 1, \dots, N$, is reasonably close to the discrete uniform pmf $\mathcal{U}\{1, 2, \dots, N\}$ then the resampling is not needed. Otherwise, the resampling is applied. Below, we show that the formula $P_2^{(N)}$ is related to the Euclidean distance between these two pdfs. It is natural to think to employ other kind of distance or discrepancy measures between the pmf represented by weights $\bar{w}_n$ and the discrete uniform pmf. We derive some alternative ESS functions below. It is important to observe that several ESS functions can be obtained from the same discrepancy measure. Here, in this section, we derive some ESS functions that satisfies some important properties, described exhaustively in the next section.

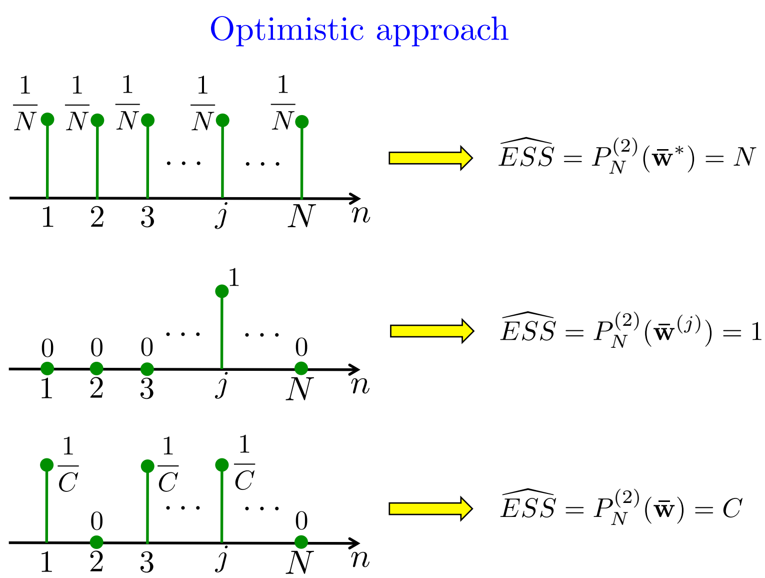


Figure 2: Graphical summary of the optimistic approach employed by $\widehat{ESS} = P_N^{(2)}(\bar{\mathbf{w}})$.

Euclidean distance L_2. Let us consider the Euclidean distance L_2 between the discrete uniform pmf $\mathcal{U}\{1, 2, \dots, N\}$ and the pmf given by the normalized weights $\bar{w}_n$, i.e,

$$\begin{aligned} L_2 &= \sqrt{\sum_{n=1}^{N} \left(\bar{w}_n - \frac{1}{N} \right)^2} \\ &= \sqrt{\left(\sum_{n=1}^{N} \bar{w}_n^2 \right) + N \left(\frac{1}{N^2} \right) - \frac{2}{N} \sum_{n=1}^{N} \bar{w}_n} \\ &= \sqrt{\left(\sum_{n=1}^{N} \bar{w}_n^2 \right) - \frac{1}{N}} \\ &= \sqrt{\frac{1}{P_N^{(2)}(\bar{\mathbf{w}})} - \frac{1}{N}}. \end{aligned} \tag{15}$$

We can observe the relationship with $P_N^{(2)}(\bar{\mathbf{w}})$. Maximizing $P_N^{(2)}$ is equivalent to minimizing the Euclidean distance between the pmf $\bar{w}_n$ and the discrete uniform pmf. Thus, the use of $P_N^{(2)}$ within a PMC or SMC scheme corresponds to check if the current pmf defined by the weights $\bar{w}_n$ is close enough to the uniform pmf, in terms of Euclidean distance.

Distance L_∞. Given the previous observations, we can attempt to obtain other suitable ESS formulas, employing other distances. Considering the distance L_∞, we have

$$\begin{aligned} L_\infty &= \max \left[\left| \bar{w}_1 - \frac{1}{N} \right|, \dots, \left| \bar{w}_N - \frac{1}{N} \right| \right] \\ &= \max \left[\bar{w}_1, \dots, \bar{w}_N \right] - \frac{1}{N} \\ &= \frac{1}{D_N^{(\infty)}(\bar{\mathbf{w}})} - \frac{1}{N}. \end{aligned} \tag{16}$$

where

$$D_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{\max \left[\bar{w}_1, \dots, \bar{w}_N \right]}, \tag{17}$$

is another valid ESS measure based on the discrepancy approach. Clearly, we have also

$$1 \leq D_N^{(\infty)}(\bar{\mathbf{w}}) \leq N,$$

with $D_N^{(\infty)}(\bar{\mathbf{w}}^*) = N$ and $D_N^{(\infty)}(\bar{\mathbf{w}}^{(i)}) = 1$ for all $i \in \{1, \dots, N\}$. Maximizing $D_N^{(\infty)}$ is equivalent to minimizing the L_∞ distance between the pmf $\bar{w}_n$ and the discrete uniform pmf.

Distance L_1. In Appendix B we show that the L_1 distance can expressed as function of a suitable ESS function $Q_N(\bar{\mathbf{w}})$, i.e.,

$$L_1 = \sum_{n=1}^{N} \left| \bar{w}_n - \frac{1}{N} \right| = 2 \left[\frac{Q_N(\bar{\mathbf{w}}) - N}{N} \right] + 2, \tag{18}$$

where

$$Q_N(\bar{\mathbf{w}}) = -N \sum_{i=1}^{N^+} \bar{w}_i^+ + N^+ + N, \tag{19}$$

with

$$\{\bar{w}_1^+, \dots, \bar{w}_{N^+}^+\} = \left\{ \text{all } \bar{w}_n: \quad \bar{w}_n \geq 1/N, \quad \forall n = 1, \dots, N \right\},$$

and $N^+ = \#\{\bar{w}_1^+, \dots, \bar{w}_{N^+}^+\}$. Note that $1 \leq Q_N(\bar{\mathbf{w}}) \leq N$, with $Q_N(\bar{\mathbf{w}}^*) = N$ and $Q_N(\bar{\mathbf{w}}^{(i)}) = 1$ for all $i \in \{1, \dots, N\}$. Maximizing Q_N is equivalent to minimizing the L_1 distance between the pmf $\bar{w}_n$ and the discrete uniform pmf. We remark again this is only one of the possible ESS functions induced by the L_1 distance. We choose this one since it satisfy some properties described in the next section.

Distance L_0. It is importance to note that interesting ESS formulas can be also obtained considering also the distance of the vector $\bar{\mathbf{w}}$ with respect to the null vector containing all zeros as entries. For instance, based on the *Hamming distance* among the two vectors, i.e.,

$$V_N^{(0)}(\bar{\mathbf{w}}) = N - N_Z, \tag{20}$$

where N_z is the number of zeros in $\bar{\mathbf{w}}$, i.e.,

$$N_Z = \#\{\bar{w}_n = 0, \quad \forall n = 1, \dots, N\}. \tag{21}$$

Observe that $1 \leq V_N^{(0)}(\bar{\mathbf{w}}) \leq N$ and $V_N^{(0)}(\bar{\mathbf{w}}^*) = N$ and $V_N^{(0)}(\bar{\mathbf{w}}^{(i)}) = 1$ for all $i \in \{1, \dots, N\}$.

All the expression $P_N^{(2)}$, Q_N, $D_N^{(\infty)}$ and $V_N^{(0)}$ share interesting properties which we list in the next section, which allow us to define a generic ESS measure based on the discrepancy between the pmf given by $\bar{\mathbf{w}}$ and the discrete uniform pmf. Other formulas related to L_p distances are obtained in Section 4.3.

4 Generalized ESS functions

In this section, we introduce some properties that a generalized ESS measure (G-ESS) should satisfy, based only on the information of the normalized weights. Here, first of all, note that any possible G-ESS is a function of the vector of normalized weights $\bar{\mathbf{w}} = [\bar{w}_1, \dots, \bar{w}_N]$,

$$E_N(\bar{\mathbf{w}}) = E_N(\bar{w}_1, \dots, \bar{w}_N) : \mathcal{S}_N \rightarrow [1, N], \tag{22}$$

where $\mathcal{S}_N \subset \mathbb{R}^N$ represents the *unit simplex* in $\mathbb{R}^N$. Namely, the variables $\bar{w}_1, \dots, \bar{w}_N$ are subjected to the constrain

$$\bar{w}_1 + \bar{w}_2 + \ldots + \bar{w}_N = 1. \tag{23}$$

4.1 Conditions for G-ESS functions

Below we list five conditions that $E_N(\bar{\mathbf{w}})$ must fulfill to be consider a suitable G-ESS function. The first three properties are strictly necessary, whereas the last two are *welcome* conditions, i.e., no strictly required but desirable (see also classification below):

C1. **Symmetry:** E_N must be invariant under any permutation of the weights, i.e.,

$$E_N(\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_N) = E_N(\bar{w}_{j_1}, \bar{w}_{j_2}, \ldots, \bar{w}_{j_N}), \tag{24}$$

for any possible set of indices $\{j_1, \ldots, j_N\} = \{1, \ldots, N\}$.

C2. **Maximum condition:** A maximum is reached at $\bar{\mathbf{w}}^*$ in Eq. (10) and has value N, i.e.,

$$E_N\left(\bar{\mathbf{w}}^*\right) = N \geq E_N(\bar{\mathbf{w}}). \tag{25}$$

C3. **Minimum condition**: the minimum value is 1 and it is reached (at least) at the vertices $\bar{\mathbf{w}}^{(j)}$ of the unit simplex in Eq. (11),

$$E_N(\bar{\mathbf{w}}^{(j)}) = 1 \leq E_N(\bar{\mathbf{w}}). \tag{26}$$

for all $j \in \{1, \ldots, N\}$.

C4. **Unicity of extreme values:** (*welcome condition*) The maximum at $\bar{\mathbf{w}}^*$ is unique and the the minimum value 1 is reached *only* at the vertices $\bar{\mathbf{w}}^{(j)}$, for all $j \in \{1, \ldots, N\}$.

C5. **Stability - Invariance of the rate $\frac{E_N(\bar{\mathbf{w}})}{N}$:** (*welcome condition*) Consider the vector of weights $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N] \in \mathbb{R}^N$ and the vector

$$\bar{\mathbf{v}} = [\bar{v}_1, \ldots, \bar{v}_{MN}] \in \mathbb{R}^{MN}, \quad M \geq 1, \tag{27}$$

obtained repeating and scaling by $\frac{1}{M}$ the entries of $\bar{\mathbf{w}}$, i.e.,

$$\bar{\mathbf{v}} = \frac{1}{M}[\underbrace{\bar{\mathbf{w}}, \bar{\mathbf{w}}, \ldots, \bar{\mathbf{w}}}_{M-times}]. \tag{28}$$

Note that, clearly, $\sum_{i=1}^{mN} \bar{v}_i = \frac{1}{M}\left[M \sum_{n=1}^{N} \bar{w}_n\right] = 1$. The invariance condition is expressed as

$$\begin{aligned} \frac{E_N(\bar{\mathbf{w}})}{N} &= \frac{E_{MN}(\bar{\mathbf{v}})}{MN} \\ E_N(\bar{\mathbf{w}}) &= \frac{1}{M} E_{MN}(\bar{\mathbf{v}}), \end{aligned} \tag{29}$$

for all $M \in \mathbb{N}^+$.

The last (welcome) condition C5 is related to the *optimistic approach* described in in Section 3. For clarifying this point, as an example, let us consider the vectors

$$\begin{aligned}
\bar{\mathbf{w}} &= [0,1,0], \\
\bar{\mathbf{v}}' &= \left[0,\frac{1}{2},0,0,\frac{1}{2},0\right] = \frac{1}{2}[\bar{\mathbf{w}},\bar{\mathbf{w}}], \\
\bar{\mathbf{v}}'' &= \left[0,\frac{1}{3},0,0,\frac{1}{3},0,0,\frac{1}{3},0\right] = \frac{1}{3}[\bar{\mathbf{w}},\bar{\mathbf{w}},\bar{\mathbf{w}}],
\end{aligned}$$

with $N = 3$. Following the optimistic approach, we should have $E_N(\bar{\mathbf{w}}) = 1$, $E_{2N}(\bar{\mathbf{v}}') = 2$ and $E_{3N}(\bar{\mathbf{v}}'') = 3$, i.e., the rate E_N/N is invariant

$$\frac{E_N(\bar{\mathbf{w}})}{N} = \frac{E_{2N}(\bar{\mathbf{v}}')}{2N} = \frac{E_{3N}(\bar{\mathbf{v}}'')}{3N} = \frac{1}{3}.$$

4.2 Classification and examples of G-ESS functions

We divide the possible G-ESS functions in different categories depending on the conditions fulfilled by the corresponding function (see Table 1). Recall that the first three conditions are strictly required. All the G-ESS functions which satisfy at least the first four conditions, i.e., from C1 to C4, are *proper* functions. All the G-ESS functions which satisfy the first three conditions, C1, C2 and C3 but no C4, are considered *degenerate* functions. When a G-ESS function fulfills the last condition is called *stable*. Thus, the G-ESS functions which satisfy all the conditions, i.e., from C1 to C5, are then *proper* and *stable* whereas, if C4 is not satisfied, they are *degenerate* and *stable*. We can also distinguish two type of degeneracy: *type-1* when $E_N(\bar{\mathbf{w}})$ reaches the maximum value N also in some other point $\bar{\mathbf{w}} \neq \bar{\mathbf{w}}^*$, or *type-2* if $E_N(\bar{\mathbf{w}})$ reaches the minimum value 1 also in some point that is not a vertex.

Table 1: Classification of G-ESS depending of the satisfied conditions.

Class of G-ESS	**C1**	**C2**	**C3**	**C4**	**C5**
Degenerate (D)	Yes	Yes	Yes	No	No
Proper (P)	Yes	Yes	Yes	Yes	No
Degenerate and Stable (DS)	Yes	Yes	Yes	No	Yes
Proper and Stable (PS)	Yes	Yes	Yes	Yes	Yes

Example 1. *The functions $P_N^{(2)}$ in Eq.(6), $D_N^{(\infty)}$ in Eq. (17) and Q_N in Eq. (19) described in Sections 3 are all G-ESS functions of class PS, proper and stable.*

Example 2. *The G-ESS function in Eq.(20), $V_N^{(0)}(\bar{\mathbf{w}}) = N - N_Z$, where N_Z is the number of zero within $\bar{\mathbf{w}}$, is degenerate (type-1) and stable.*

Example 3. *Let us denote the harmonic mean of the normalized weights as*

$$\mathrm{HarM}(\bar{\mathbf{w}}) = \frac{1}{\sum_{n=1}^{N} \frac{1}{\bar{w}_n}},$$

Note that $\lim_{\bar{w}_n \to 0} \mathrm{HarM}(\bar{\mathbf{w}}) = 0$ *for any possible entry in* $\bar{\mathbf{w}}$*, i.e.,* $\forall n \in \{1, \ldots, N\}$*. The following functions involving the harmonic mean,*

$$A_{1,N}(\bar{\mathbf{w}}) = \frac{1}{(1-N)\mathrm{HarM}(\bar{\mathbf{w}}) + 1}, \quad (30)$$

$$A_{2,N}(\bar{\mathbf{w}}) = (N^2 - N)\mathrm{HarM}(\bar{\mathbf{w}}) + 1, \quad (31)$$

are both degenerate (type-2) G-ESS functions. They satisfy C1, C2 and C3, whereas C4 and C5 are not fulfilled.

Example 4. *The following functions involving the minimum of the normalized weights,*

$$T_{1,N}(\bar{\mathbf{w}}) = \frac{1}{(1-N)\min[\bar{w}_1, \ldots, \bar{w}_N] + 1}, \quad (32)$$

$$T_{2,N}(\bar{\mathbf{w}}) = (N^2 - N)\min[\bar{w}_1, \ldots, \bar{w}_N] + 1, \quad (33)$$

are degenerate (type-2) G-ESS measures.

Example 5. *The perplexity function introduced in [1], is defined as*[2]

$$\mathrm{Per}_N(\bar{\mathbf{w}}) = 2^{H(\bar{\mathbf{w}})}, \quad (34)$$

where

$$H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log_2(\bar{w}_n), \quad (35)$$

is the discrete entropy [5] of the pmf $\bar{w}_n$*,* $n = 1, \ldots, N$*. The perplexity is a proper and stable G-ESS function.*

Example 6. *Let us consider the Gini coefficient* $G(\bar{\mathbf{w}})$ *[14, 20], defined as follows. First of all, we define the non-decreasing sequence of normalized weights*

$$\bar{w}_{(1)} \leq \bar{w}_{(2)} \leq \ldots \leq \bar{w}_{(N)}, \quad (36)$$

obtained sorting in ascending order the entries of the vector $\bar{\mathbf{w}}$*. The Gini coefficient is defined as*

$$G(\bar{\mathbf{w}}) = 2\frac{s(\bar{\mathbf{w}})}{N} - \frac{N+1}{N}, \quad (37)$$

where

$$s(\bar{\mathbf{w}}) = \sum_{n=1}^{N} n\bar{w}_{(n)}. \quad (38)$$

[2] We have slightly modified the definition of the perplexity for fitting better in our framework.

Then, the G-ESS function defined as

$$\text{Gini}_N(\bar{\mathbf{w}}) = -NG(\bar{\mathbf{w}}) + N, \tag{39}$$

is proper and stable.

Example 7. *The following G-ESS function (inspired by the L_1 distance),*

$$\text{N-plus}_N(\bar{\mathbf{w}}) = N^+ = \#\big\{\bar{w}_n \geq 1/N, \quad \forall n = 1, \dots, N\big\}. \tag{40}$$

is also degenerate (type 2) and stable.

4.3 Design of G-ESS families

We can easily design G-ESS functions fulfilling at least the first three conditions, C1, C2, and C3. As examples, considering a parameter $r \geq 0$, we introduce four families of G-ESS functions which have the following analytic forms

$$P_N^{(r)}(\bar{\mathbf{w}}) = \frac{1}{a_r \sum_{n=1}^N (\bar{w}_n)^r + b_r}, \quad r \in \mathbb{R} \quad D_N^{(r)}(\bar{\mathbf{w}}) = \frac{1}{a_r \left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{\frac{1}{r}} + b_r}, \quad r \geq 0$$

$$V_N^{(r)}(\bar{\mathbf{w}}) = a_r \sum_{n=1}^N (\bar{w}_n)^r + b_r, \quad r \in \mathbb{R} \qquad S_N^{(r)}(\bar{\mathbf{w}}) = a_r \left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{\frac{1}{r}} + b_r, \qquad r \geq 0$$

where a_r, b_r are constant values depending on the parameter r. The values of the coefficients a_r, b_r can be found easily as solutions of linear systems (see Appendix A), with equations obtained in order to fulfill the conditions C2 and C3. The resulting G-ESS functions are in general *proper*, i.e., satisfying from C1 to C4 (with some degenerate and stable exceptions). The solutions of the corresponding linear systems are given in Table 2. Replacing these solutions within the expressions of the different families, we obtain

$$P_N^{(r)}(\bar{\mathbf{w}}) = \frac{N^{(2-r)} - N}{(1-N)\sum_{n=1}^N (\bar{w}_n)^r + N^{(2-r)} - 1}, \tag{41}$$

$$D_N^{(r)}(\bar{\mathbf{w}}) = \frac{N^{\frac{1}{r}} - N}{(1-N)\left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{\frac{1}{r}} + N^{\frac{1}{r}} - 1}, \tag{42}$$

$$V_N^{(r)}(\bar{\mathbf{w}}) = \frac{N^{r-1}(N-1)}{1-N^{r-1}}\left[\sum_{n=1}^N \bar{w}_n^r\right] + \frac{N^r - 1}{N^{r-1} - 1}, \tag{43}$$

$$S_N^{(r)}(\bar{\mathbf{w}}) = \frac{N-1}{N^{\frac{1-r}{r}} - 1}\left[\left(\sum_{n=1}^N \bar{w}_n^r\right)^{\frac{1}{r}}\right] + 1 - \frac{N-1}{N^{\frac{1-r}{r}} - 1}, \tag{44}$$

These families contain different G-ESS functions previously introduced, and also other interesting special cases. Table 5 summarizes these particular cases (jointly with the corresponding classification) corresponding to specific values the parameter r. Some of them involve the *geometric mean* of the normalized weights,

$$\text{GeoM}(\bar{\mathbf{w}}) = \left[\prod_{n=1}^{N} \bar{w}_n\right]^{1/N}, \tag{45}$$

other ones involve the *discrete entropy* [5] of the normalized weights,

$$H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log_2(\bar{w}_n), \tag{46}$$

and others use the number of zeros contained in $\bar{\mathbf{w}}$, $N_Z = \#\{\bar{w}_n = 0, \quad \forall n = 1, \ldots, N\}$. The derivations of these special cases are provided in Appendices A.1 and A.2. Note that Table 5 contains a proper and stable G-ESS function

$$S_N^{(1/2)}(\bar{\mathbf{w}}) = \left(\sum_{n=1}^{N} \sqrt{\bar{w}_n}\right)^2, \tag{47}$$

not introduced so far. Figure 7 shows the rate $\frac{1}{N} \leq \frac{ESS}{N} \leq 1$ as function of r for the different families $P_N^{(r)}$, $D_N^{(r)}$ (both in solid lines), $V_N^{(r)}$ and $S_N^{(r)}$ (both in dashed lines), considering the vector $\bar{\mathbf{w}} = [0.1\ 0.1\ 0.2\ 0.6]$ in (a) $\bar{\mathbf{w}} = [0\ 0\ 0.5\ 0.5]$ in (c) and its repeating versions in (b)-(d) (i.e., consider the vector $\bar{\mathbf{v}}$ obtained repeating $M = 100$ times the entries of $\bar{\mathbf{w}}$ and scaling by $1/M$). Note that $V_N^{(0)}/N$, $P_N^{(2)}/N$, $S_N^{(\frac{1}{2})}/N$ and $D_N^{(\infty)}/N$ remain invariant, in both cases. In (c)-(d), since the vector $\bar{\mathbf{w}} = [0\ 0\ 0.5\ 0.5]$ contains two zeros, the rates $V_N^{(0)}/N$, $P_N^{(2)}/N$, $S_N^{(\frac{1}{2})}/N$, $D_N^{(\infty)}/N$ remain invariant equal to the same value 0.5, in this case. This means that the corresponding ESS values are 2 and 200, as expected.

Table 2: G-ESS families and their coefficients a_r and b_r.

$P_N^{(r)}(\bar{\mathbf{w}})$	$D_N^{(r)}(\bar{\mathbf{w}})$	$V_N^{(r)}(\bar{\mathbf{w}})$	$S_N^{(r)}(\bar{\mathbf{w}})$
$\frac{1}{a_r \sum_{n=1}^{N} (\bar{w}_n)^r + b_r}$	$\frac{1}{a_r \left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{\frac{1}{r}} + b_r}$	$a_r \sum_{n=1}^{N} (\bar{w}_n)^r + b_r$	$a_r \left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{\frac{1}{r}} + b_r$
$P_N^{(r)}(\bar{\mathbf{w}})$	$D_N^{(r)}(\bar{\mathbf{w}})$	$V_N^{(r)}(\bar{\mathbf{w}})$	$S_N^{(r)}(\bar{\mathbf{w}})$
$a_r = \frac{1-N}{N^{(2-r)}-N}$	$a_r = \frac{N-1}{N-N^{\frac{1}{r}}}$	$a_r = \frac{N^{r-1}(N-1)}{1-N^{r-1}}$	$a_r = \frac{N-1}{N^{\frac{1-r}{r}}-1}$
$b_r = \frac{N^{(2-r)}-1}{N^{(2-r)}-N}$	$b_r = \frac{1-N^{\frac{1}{r}}}{N-N^{\frac{1}{r}}}$	$b_r = \frac{N^r-1}{N^{r-1}-1}$	$b_r = \frac{N^{\frac{1-r}{r}}-N}{N^{\frac{1-r}{r}}-1}$

4.4 Distribution of the ESS values

An additional feature of the G-ESS measures is related to the distribution of the effective sample size values obtained with a specific G-ESS function, when the vector $\bar{\mathbf{w}}$ is considered as a realization of a random variable uniformly distributed in the unit simplex $\mathcal{S}_N$. Namely, let us consider the random variables $\bar{\mathbf{W}} \sim \mathcal{U}(\mathcal{S}_N)$ and $E = E_N\left(\bar{\mathbf{W}}\right)$ with probability density function (pdf) $p_N(e)$, i.e.,

$$E = E_N\left(\bar{\mathbf{W}}\right) \sim p_N(e). \tag{48}$$

Clearly, the support of $p_N(e)$ is $[1, N]$. Studying $p_N(e)$, we can define additional properties for discriminating different G-ESS functions. For instance, in general $p_N(e)$ is not a uniform pdf. Some functions E_N concentrate more probability mass closer to the extreme N, other functions closer to the extreme 1. This feature varies with N, in general. For $N = 2$, it is straightforward to obtain the expression of the pdf $p_2(e)$ for certain G-ESS functions. Indeed, denoting as $I_1(e)$ and $I_2(e)$ the inverse functions corresponding to the monotonic pieces of the generic function $E_2\left(\bar{w}_1, 1 - \bar{w}_1\right) = E_2(\bar{w}_1)$, then we obtain

$$p_2(e) = \left|\frac{dI_1}{de}\right| + \left|\frac{dI_2}{de}\right|, \qquad e \in [1, N], \tag{49}$$

using the expression of transformation of a uniform random variable, defined in $[0, 1]$. Thus, we find that $p_2(e) = \frac{2}{e^2}$ for $D_2^{(\infty)}$ and $p_2(e) = \frac{2}{e^2\sqrt{\frac{2}{e}-1}}$ for $P_2^{(2)}$, for instance. Figure 6 depicts the pdfs $p_2(e)$ for $D_2^{(\infty)}$ in Eq. (17), $T_{2,2}$ in Eq. (33) and $P_2^{(2)}$ in Eq. (6). We can observe that $P_2^{(2)}$ is *more optimistic* than $D_2^{(\infty)}$ *judging* a set of weighted samples and assigning a value of the effective size, since $p_2(e)$ in this case is unbalanced to the right side close to 2. From a practical point of view, the pdf $p_N(e)$ could be used for choosing the threshold values for the adaptive resampling. The *limiting distribution* obtained for $N \to \infty$,

$$p_\infty(e) = \lim_{N\to\infty} p_N(e), \tag{50}$$

is also theoretically interesting, since it can characterize the function E_N. However, it is not straightforward to obtain $p_\infty(e)$ analytically. In Section 5, we approximate different limiting pdfs $p_\infty(e)$ of different G-ESS functions via numerical simulation.

4.5 Summary

In the previous sections, we have found different *stable* G-ESS functions, satisfying at least the conditions C1, C2, C3, and C5. They are recalled in Table 3. The following ordering inequalities

$$D_N^{(\infty)}(\bar{\mathbf{w}}) \leq P_N^{(2)}(\bar{\mathbf{w}}) \leq S_N^{(\frac{1}{2})}(\bar{\mathbf{w}}) \leq V_N^{(0)}(\bar{\mathbf{w}}), \quad \forall \bar{\mathbf{w}} \in \mathcal{S}_N,$$

can be also easily proved. For the case $N = 2$, i.e., when we have $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2]$ with $\bar{w}_2 = 1 - \bar{w}_1$, several interesting relationships can be found as shown in Appendix C. For instance, the standard formula $P_2^{(2)}$ for $N = 2$ is identical to the G-ESS function in Eq. (30) involving the harmonic mean.

Table 3: Stable G-ESS functions.

$D_N^{(\infty)}(\bar{\mathbf{w}})$	$P_N^{(2)}(\bar{\mathbf{w}})$	$S_N^{(\frac{1}{2})}(\bar{\mathbf{w}})$	$V_N^{(0)}(\bar{\mathbf{w}})$
$\frac{1}{\max[\bar{w}_1,\ldots,\bar{w}_N]}$	$\frac{1}{\sum_{n=1}^N \bar{w}_n^2}$	$\left(\sum_{n=1}^N \sqrt{\bar{w}_n}\right)^2$	$N - N_Z$
proper	*proper*	*proper*	*degenerate (type-1)*

$Q_N(\bar{\mathbf{w}})$	$\text{N-plus}_N(\bar{\mathbf{w}})$	$\text{Gini}_N(\bar{\mathbf{w}})$	$\text{Per}_N(\bar{\mathbf{w}})$
$-N\sum_{i=1}^{N^+} \bar{w}_i^+ + N^+ + N$	N^+	$-NG(\bar{\mathbf{w}}) + N$	$2^{H(\bar{\mathbf{w}})}$
proper	*degenerate (type-2)*	*proper*	*proper*

5 Simulations

5.1 Analysis of the distribution of ESS values

In this section, we study the distribution $p_N(e)$ of the values of the different G-ESS families. With this purpose, we draw different vectors $\bar{\mathbf{w}}'$ uniformly distributed in the unit simplex $\mathcal{S}_N$, and then we compute the corresponding ESS values. For drawing uniformly in the standard simplex $\mathcal{S}_N \subset \mathbb{R}^N$, we use the following procedure involving *uniform spacings* random variables [6]:

1. Draw N uniform random samples $u_n \sim \mathcal{U}([0,1])$, $n = 1, \ldots, N$ and sort them in ascending order obtaining $u_{(1)} < u_{(2)} < \ldots < u_{(N)}$.

2. Set $s_1 = u_{(1)}$, $s_2 = u_{(2)} - u_{(1)}$, $\ldots$, $s_{N-1} = u_{(N)} - u_{(N-1)}$, and $s_N = 1 - u_{(N)}$. Note that $\sum_{n=1}^N s_n = 1$.

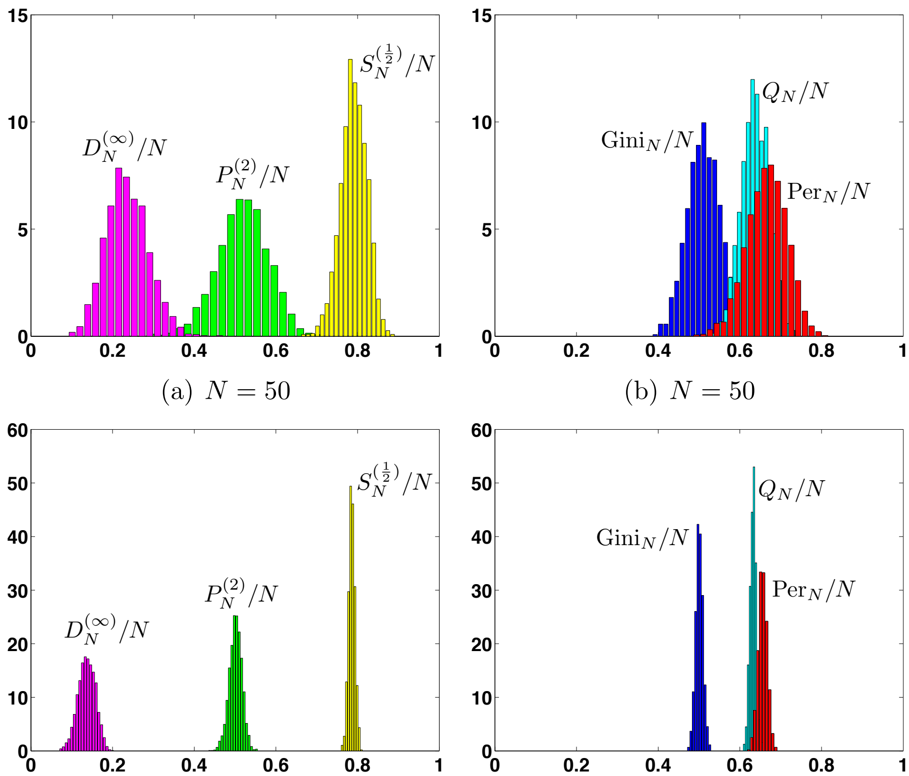
3. Given the N vertices $\bar{\mathbf{w}}^{(j)} = [0, \ \ldots 0, \ 1, \ 0, \ldots, 0]$ of the unit simplex, $j = 1, \ldots, N$, then a random vector uniform distributed in the simplex is defined as

$$\bar{\mathbf{w}}' = \sum_{n=1}^N s_n \bar{\mathbf{w}}^{(n)}. \tag{51}$$

We repeat the procedure above 2000 times (for each value of N), namely, we generate 2000 independent random vectors $\bar{\mathbf{w}}'$ uniformly distributed in the unit simplex $\mathcal{S}_N \subset \mathbb{R}^N$. After that we evaluate the different proper and stable G-ESS functions (summarized in Table 3) at each drawn vector $\bar{\mathbf{w}}'$. The resulting histograms of the rate ESS/N obtained by the different functions are depicted in Figure 3. Figures 3(a)-(c) correspond to $N = 50$, whereas (b)-(d) correspond to $N = 1000$. Figures 3(a)-(c) show the histograms of the rate corresponding $D_N^{(\infty)}$, $P_N^{(2)}$, $S_N^{(\frac{1}{2})}$, whereas Figures (b)-(d) show the histograms of the rate corresponding Q_N, Gini_N and Per_N. The empirical means and standard deviations for different N are provided in Table 4.

Table 4: Statistics of $\widehat{p}_N(e)$, empirical approximation of $p_N(e)$, corresponding to different G-ESS functions. The greatest standard deviations for a given N are highlighted with boldface.

Description	**N**	$\mathbf{D_N^{(\infty)}/N}$	$\mathbf{P_N^{(2)}/N}$	$\mathbf{S_N^{(\frac{1}{2})}/N}$	$\mathbf{Q_N/N}$	$\mathbf{Gini_N/N}$	$\mathbf{Per_N/N}$
mean	50	0.2356	0.5194	0.7902	0.6371	0.5117	0.6655
	200	0.1776	0.5057	0.7868	0.6326	0.5020	0.6568
	10^3	0.1366	0.5013	0.7858	0.6324	0.5007	0.6558
	$5\ 10^3$	0.1121	0.5005	0.7856	0.6322	0.5002	0.6554
std	50	0.0517	**0.0622**	0.0324	0.0345	0.0410	0.0492
	200	0.0336	**0.0341**	0.0168	0.0171	0.0204	0.0248
	10^3	**0.0213**	0.0158	0.0077	0.0077	0.0091	0.0111
	$5\ 10^3$	**0.0145**	0.0071	0.0034	0.0034	0.0040	0.0050



(a) $N = 50$ (b) $N = 50$

(c) $N = 1000$ (d) $N = 1000$

Figure 3: The histograms of the rates $\frac{ESS}{N}$ corresponding to the proper and stable G-ESS functions in Table 3, with $N \in \{50, 1000\}$.

We can observe that all the G-ESS functions concentrate the probability mass of the ESS values around one mode, located in different positions. The variances of these distributions decrease as N grows. The statistical information provided by these histograms can be used for choosing the threshold value in an adaptive resampling scheme. Typically, the condition for applying resampling is

$$E_N(\bar{\mathbf{w}}) \leq \epsilon N,$$

where $0 \leq \epsilon \leq 1$. Namely, the information provided by Table 4 can be useful for choosing ϵ, depending on the used G-ESS function. For instance, Doucet et al. [10, Section 3.5] suggest to use $\epsilon = \frac{1}{2}$ for $P_N^{(2)}$. This suggestion can be explained considering the mean of the ESS values of $P_N^{(2)}$, which is ≈ 0.5. Moreover, the standard deviation can help us to understand the capability of each formula in differentiating different vectors $\bar{\mathbf{w}}$. The greatest standard deviation for each N is highlighted with boldface. In this sense, $D_N^{(\infty)}$ seems the most "discriminative" for large values of N, whereas $P_N^{(2)}$ seems the more convenient for small values of N (however, other studies can suggest the opposite; see below).

5.2 Approximation of the theoretical ESS definition

Let us recall the theoretical definition of ESS in Eq. (5),

$$ESS_{var} = N \frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}. \tag{52}$$

We have already discussed in Section 2.1, a more convenient definition is

$$ESS_{MSE} = N \frac{\text{MSE}_\pi[\widehat{I}]}{\text{MSE}_q[\widetilde{I}]} = N \frac{\text{var}_\pi[\widehat{I}]}{\text{MSE}_q[\widetilde{I}]}. \tag{53}$$

considering the Mean Square Error (MSE) of the estimators, instead of only the variance. For large values of N the difference between the two definitions is negligible since the bias of $\widetilde{I}$ is virtually zero. In this section, we compute approximately via Monte Carlo the theoretical definitions ESS_{var}, ESS_{MSE}, and compare with the values obtained with different G-ESS functions. More specifically, we consider a univariate standard Gaussian density as target pdf,

$$\bar{\pi}(x) = \mathcal{N}(x; 0, 1), \tag{54}$$

and alos a Gaussian proposal pdf,

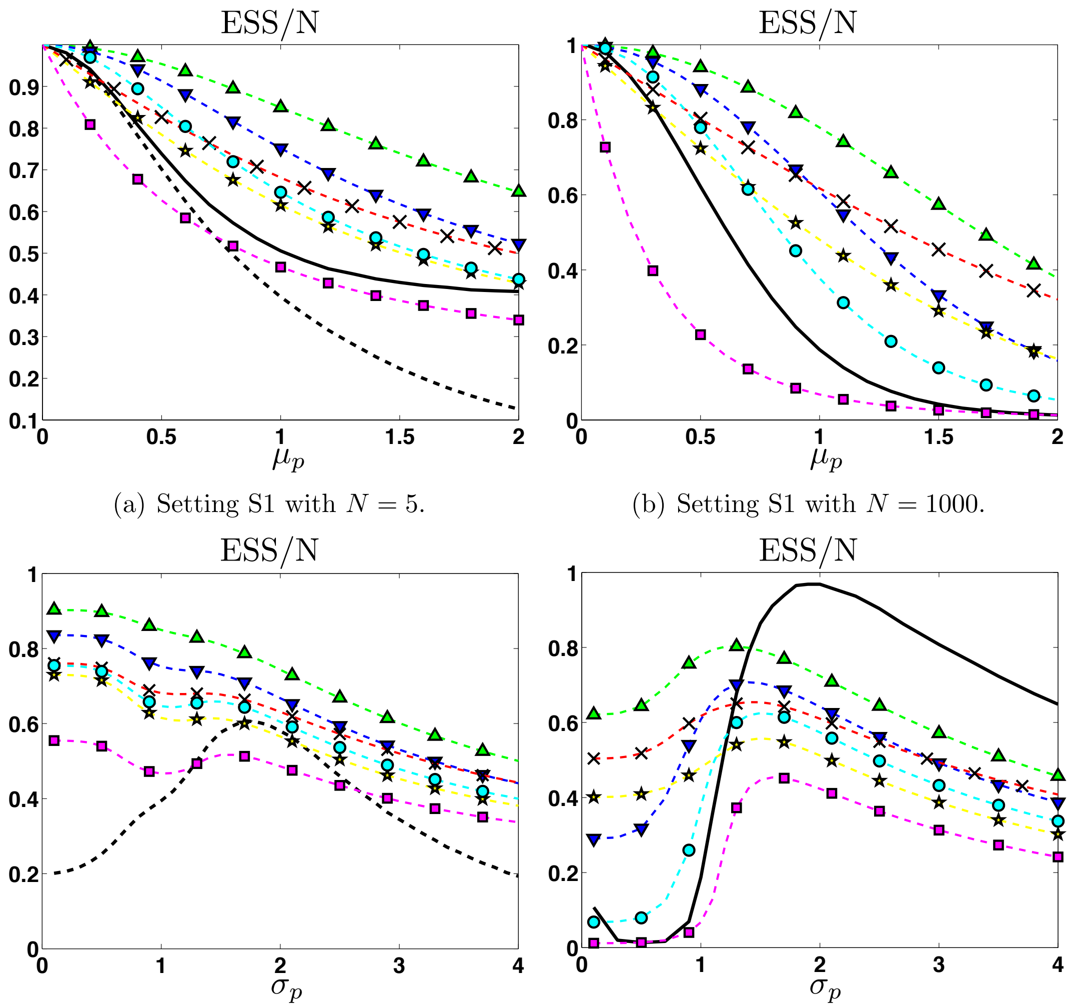
$$q(x) = \mathcal{N}(x; \mu_p, \sigma_p^2), \tag{55}$$

with mean μ_p and variance σ_p^2. Furthermore, we consider two experiment settings:

S1 In this scenario, we set $\sigma_p = 1$ and vary $\mu_p \in [0, 2]$. Clearly, for $\mu_p = 0$ we have the ideal Monte Carlo case, $q(x) \equiv \bar{\pi}(x)$. As μ_p increases, the proposal becomes more different from $\bar{\pi}$.

S2 In this case, we set $\mu_p = 1$ and consider $\sigma_p \in [0.1, 4]$.

In both cases, we test $N \in \{5, 1000\}$. Figure 4 shows the (approximated) theoretical ESS curves and the curves corresponding to the proper and stable G-ESS formulas. For $N = 1000$, the difference between ESS_{var} and ESS_{MSE} is negligible, so that we only show ESS_{var}. For $N = 5$ and S1 we show both curves of ESS_{var} and ESS_{MSE}, whereas for $N = 5$ and S2 we only provide ESS_{MSE} since the bias is big for small value of σ_p so that it is difficult to obtain reasonable and meaningful values of ESS_{var}.



(a) Setting S1 with $N = 5$.

(b) Setting S1 with $N = 1000$.

(c) Setting S2 with $N = 5$.

(d) Setting S2 with $N = 1000$.

Figure 4: Rates corresponding to ESS_{var} (solid line), ESS_{MSE} (dashed line; shown only in (a)-(c)), $P_N^{(2)}$ (circles), $D_N^{(\infty)}$ (squares), Gini_N (stars), $S_N^{(1/2)}$ (triangles up), Q_N (x-marks), Per_N (triangles down).

In the setting S1 with $N = 5$ shown Fig. 4(a), first of all we can observe that ESS_{var} and ESS_{MSE} are very close when $\mu_p \approx 0$ (i.e., $q(x) \approx \bar{\pi}(x)$) but they differ substantially when the bias increases. In this case, the G-ESS function Gini_N provides the closest values to ESS_{var}, in general. Moreover, $P_N^{(2)}$ and $D_N^{(\infty)}$ also provide good approximations of ESS_{var}. Note that ESS_{var} is always contained between $D_N^{(\infty)}$ and $P_N^{(2)}$. In the case S1 with $N = 1000$ shown Fig.

4(b), the formula $P_N^{(2)}$ provides the closest curve to ESS_{var}. The G-ESS function $D_N^{(\infty)}$ gives a good approximation when μ_p increases, i.e., the scenario becomes worse from a Monte Carlo point of view. The G-ESS function Gini_N provides the best approximation when $\mu_p \in [0, 0.5]$. Again, ESS_{var} is always contained between $D_N^{(\infty)}$ and $P_N^{(2)}$.

In the second scenario S2 with $N = 5$ shown Fig. 4(c), all G-ESS functions are not able to reproduce conveniently the shape of ESS_{MSE}. Around to the optimal value of σ_p, Gini_N and $P_N^{(2)}$ provide the best approximation of ESS_{MSE}. For the rest of value of σ_p, $D_N^{(\infty)}$ provides the closest results. In the second setting S2 with $N = 1000$ shown Fig. 4(d), $P_N^{(2)}$ seems to emulate better the evolution of ESS_{var}. However, $D_N^{(\infty)}$ provides the closest results for small values of σ_p.

We can conclude that in general when the proposal differs substantially from the target, $D_N^{(\infty)}$ provides the best results, whereas in better scenarios and large N, $P_N^{(2)}$ seems to be the best approximations. When the proposal is quite close to the target, the function Gini_N provides also good results. The E-SS function Gini_N seems to perform better than $P_N^{(2)}$ when the number of particles N is small.

6 Conclusions

In this work, we have derived several alternative ESS functions for importance sampling. The novel ESS expressions (jointly with the formulas already presented in literature) have been classified according to five theoretical requirements presented and discussed in this work. This classification has allowed to select six different ESS functions which satisfy all these conditions. Then, we have tested them by numerical simulations. At least one of them, $D_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{\max[\bar{w}_1, \ldots, \bar{w}_N]}$ presents interesting features and some benefit, compared to the standard ESS formula $P_N^{(2)}(\bar{\mathbf{w}})$, when the proposal function differs substantially from the target distribution. Moreover, $D_N^{(\infty)}(\bar{\mathbf{w}})$ seems to behave as a "lower bound" for the theoretical ESS definition in our simulations. The simulation study also provides some useful value for choosing the threshold in an adaptive resampling context. For instance, the results in Table 4 suggest to use $\epsilon = \frac{1}{2}$ for $P_N^{(2)}$ in the resampling condition $P_N^{(2)} \leq \epsilon N$ (as already noted in [10, Section 3.5]), and $\epsilon = 0.11$ for $D_N^{(\infty)}$.

Acknowledgements

The authors would to thank Rafael Bassi Stern (UFSCAR) for some useful comments provided about this topic. This work has been supported by the Grant 2014/23160-6 of São Paulo Research Foundation (FAPESP) and by the Grant 305361/2013-3 of National Council for Scientific and Technological Development (CNPq).

References

[1] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert. Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18:447–459, 2008.

[2] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

[3] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, 2002.

[4] S. Chatterjee and P. Diaconis. The sample size required in importance sampling. *arXiv:1511.01437*, 2015.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York (USA), 1991.

[6] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.

[7] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, September 2003.

[8] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York (USA), 2001.

[9] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[10] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *technical report*, 2008.

[11] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo. Generalized multiple importance sampling. *arXiv:1511.03095*, 2015.

[12] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo. Improving population monte carlo: Alternative weighting and resampling schemes. *viXra:1601.0174*, 2015.

[13] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,*. Chapman & Hall/CRC Texts in Statistical Science, 2006.

[14] C. Gini. Measurement of inequality and incomes. *The Economic Journal*, 31:124–126, 1921.

[15] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings-F Radar and Signal Processing*, 140:107–113, 1993.

[16] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover Publications, New York, 1975.

[17] A. Kong. A note on importance sampling using standardized weights. *Technical Report 348, Department of Statistics, University of Chicago*, 1992.

[18] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

[19] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2004.

[20] M. O. Lorenz. Methods of measuring the Concentration of Wealth. *Publications of the American Statistical Association*, 9(70):209–219, 1905.

[21] L. Martino, J. Read, V. Elvira, and F. Louzada. Cooperative parallel particle filters for on-line model selection and applications to urban mobility. *viXra:1512.0420*, 2015.

[22] R. E. Megginson. *An Introduction to Banach Space Theory*. Springer, 1998.

[23] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[24] C. P. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer, 2010.

A Derivation of alternative ESS functions

It is possible to design proper G-ESS fulfilling at least the conditions C1, C2, C3 and C4 (with some degenerate exception), given in the previous section. Below, we show a possible simple procedure but several could be used. Let us consider a function $f(\bar{\mathbf{w}}) : \mathbb{R}^N \to \mathbb{R}$, which satisfies the following properties:

1. $f(\bar{\mathbf{w}})$ is a quasi-concave or a quasi-convex function, with a minimum or a maximum (respectively) at $\bar{\mathbf{w}}^* = \left[\frac{1}{N}, \dots, \frac{1}{N}\right]$.
2. $f(\bar{\mathbf{w}})$ is symmetric in the sense of Eq. (24).
3. Considering the vertices of the unit simplex $\bar{\mathbf{w}}^{(i)} = \delta(i)$ in Eq. (11), then we also assume

$$f(\bar{\mathbf{w}}^{(i)}) = c,$$

where $c \in \mathbb{R}$ is a constant value, the same for all $i = 1, \dots, N$.

Let also consider the function $af(\bar{\mathbf{w}})+b$ obtained as a linear transformation of $f(\bar{\mathbf{w}})$ where $a, b \in \mathbb{R}$ are two constants. Note that, we can always set $a > 0$ if $f(\bar{\mathbf{w}})$ is quasi-concave, or $a < 0$ if $f(\bar{\mathbf{w}})$ is quasi-convex, in order to obtain $g(\bar{\mathbf{w}})$ is always quasi-concave. Hence, we can define the G-ESS function as

$$E_N(\bar{\mathbf{w}}) = \frac{1}{af(\bar{\mathbf{w}}) + b}, \quad \text{or} \quad E_N(\bar{\mathbf{w}}) = af(\bar{\mathbf{w}}) + b, \tag{56}$$

In order to fulfill the properties 2 and 3 in Section 4, recalling $\bar{\mathbf{w}}^* = [\frac{1}{N}, \dots, \frac{1}{N}]$ and $\bar{\mathbf{w}}^{(i)} = \delta(i)$, we can properly choose the constant values a, b in order to satisfy the following system of $N+1$ equations and two unknowns a and b,

$$\begin{cases} af(\bar{\mathbf{w}}^*) + b = \dfrac{1}{N}, \\ af(\bar{\mathbf{w}}^{(i)}) + b = 1, \quad \forall i \in \{1, \dots, N\}. \end{cases} \tag{57}$$

or

$$\begin{cases} af(\bar{\mathbf{w}}^*) + b = N, \\ af(\bar{\mathbf{w}}^{(i)}) + b = 1, \quad \forall i \in \{1, \dots, N\}, \end{cases} \tag{58}$$

respectively. Note that they are both linear with respect to with unknowns a and b. Moreover, since $f(\bar{\mathbf{w}}^{(i)}) = c$ for all $i \in \{1, \dots, N\}$, the system above is reduced to a 2×2 linear system with solution

$$\begin{cases} a = \dfrac{N-1}{N[f(\bar{\mathbf{w}}^{(i)}) - f(\bar{\mathbf{w}}^*)]}, \\ b = \dfrac{f(\bar{\mathbf{w}}^{(i)}) - Nf(\bar{\mathbf{w}}^*)}{N[f(\bar{\mathbf{w}}^{(i)}) - f(\bar{\mathbf{w}}^*)]}. \end{cases} \tag{59}$$

and

$$\begin{cases} a = \dfrac{N-1}{f(\bar{\mathbf{w}}^*) - f(\bar{\mathbf{w}}^{(i)})}, \\ b = \dfrac{f(\bar{\mathbf{w}}^*) - Nf(\bar{\mathbf{w}}^{(i)})}{(\bar{\mathbf{w}}^{(i)}) - f(\bar{\mathbf{w}}^{(i)})}. \end{cases} \tag{60}$$

respectively. Below, we derive some special cases of the families $P_N^{(r)}$, $D_N^{(r)}$, $V_N^{(r)}$, and $S_N^{(r)}$ defined in Section 4.3 and obtained used the procedure above. In these families, we have $f_r(\bar{\mathbf{w}}) = \sum_{n=1}^N (\bar{w}_n)^r$ for $P_N^{(r)}$, and $V_N^{(r)}$, and $f_r(\bar{\mathbf{w}}) = \left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{1/r}$ for $D_N^{(r)}$ and $S_N^{(r)}$.

A.1 Special cases of $P_N^{(r)}(\bar{\mathbf{w}})$

In the following, we analyze some special cases of the family $P_N^{(r)}(\bar{\mathbf{w}})$ in Eq. (41):

Case $r \to 0$. In this case, the constants in Table 2 reach the values $a_r \to a_0 = -\frac{1}{N}$ and $b_r \to b_0 = \frac{N+1}{N}$. Let us define $0^0 = 0$, considering that $\lim_{r \to 0^+} 0^r = 0$ (i.e., r approaches 0 from the right). With this assumption, Thus, if no zeros are contained in $\bar{\mathbf{w}}$ then $f_0(\bar{\mathbf{w}}) = N$ and $P_N^{(0)}(\bar{\mathbf{w}}) = \frac{1}{Na_0+b_0} = N$, whereas if $\bar{\mathbf{w}}$ contains N_Z zeros, we have $f_0(\bar{\mathbf{w}}) = N - N_Z$ and

$$P_N^{(0)}(\bar{\mathbf{w}}) = \frac{N}{N_Z + 1}, \tag{61}$$

where we recall that N_Z is the number of zero within $\bar{\mathbf{w}}$. Note that, clearly, $P_N^{(0)}(\bar{\mathbf{w}}^{(i)}) = 1$ for all $i \in \{1, \dots, N\}$, since $N_Z = N - 1$.

Case $r = 1$. In this case, $a_r \to \pm\infty$, $b_r \to \mp\infty$, when $r \to 1$. Since $f_r(\bar{\mathbf{w}}) \to 1$ if $r \to 1$, we have an indeterminate form for $g_r(\bar{\mathbf{w}}) = a_r + b_r$ of type $\infty - \infty$. Note that the limit

$$\lim_{r \to 1} P_N^{(r)}(\bar{\mathbf{w}}) = \lim_{r \to 1} \frac{N^{(2-r)} - N}{(1-N)\sum_{n=1}^N (\bar{w}_n)^r + N^{(2-r)} - 1},$$

presents an indeterminate form of type $\frac{0}{0}$. Hence, using the L'Hôpital's rule [16], i.e., deriving both numerator and denominator w.r.t. r and computing the limit, we obtain

$$\begin{aligned}
P_N^{(1)}(\bar{\mathbf{w}}) &= \lim_{r\to 1} \frac{-N^{(2-r)}\log(N)}{-N^{(2-r)}\log(N) - (N-1)\sum_{n=1}^N \bar{w}_n^r \log(\bar{w}_n)}, \\
&= \frac{-N\log(N)}{-N\log(N) - (N-1)\sum_{n=1}^N \bar{w}_n \log(\bar{w}_n)}, \\
&= \frac{-N\frac{\log_2(N)}{\log_2 e}}{-N\frac{\log_2(N)}{\log_2 e} - (N-1)\sum_{n=1}^N \bar{w}_n \frac{\log_2(\bar{w}_n)}{\log_2 e}}, \\
&= \frac{-N\log_2(N)}{-N\log_2(N) + (N-1)H(\bar{\mathbf{w}})},
\end{aligned} \tag{62}$$

where we have denoted as $H(\bar{\mathbf{w}}) = -\sum_{n=1}^N \bar{w}_n \log_2(\bar{w}_n)$ the discrete entropy of the pmf $\bar{w}_n$, $n = 1, \ldots, N$. Observe that $H(\bar{\mathbf{w}}^*) = \log_2 N$ then $P_N^{(1)}(\bar{\mathbf{w}}) = \frac{-N\log_2(N)}{-\log_2 N} = N$, whereas $H(\bar{\mathbf{w}}^{(i)}) = 0$ (considering $0\log_2 0 = 0$), $P_N^{(1)}(\bar{\mathbf{w}}) = 1$.

Case $\boldsymbol{r}$= 2. In this case, $a_2 = 1$ and $b_2 = 0$, hence we obtain

$$P_N^{(2)}(\bar{\mathbf{w}}) = \frac{1}{\sum_{n=1}^N (\bar{w}_n)^2}.$$

Case $\boldsymbol{r}\to\infty$. We have $a_r \to a_\infty = \frac{N-1}{N}$ and $b_r \to b_\infty = \frac{1}{N}$. If $\bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}$ for all possible $i \in \{1, \ldots, N\}$, then we have $\lim_{r\to\infty} f_r(\bar{\mathbf{w}}) = 0$ (since $0 < \bar{w}_n < 1$, in this case) and $P_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{b_r} = N$. Otherwise, if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}$ for some $i \in \{1, \ldots, N\}$, then $\lim_{r\to\infty} f_r(\bar{\mathbf{w}}) = 1$ (where we have considered $\lim_{r\to\infty} 0^r = 0$ and $\lim_{r\to\infty} 1^r = 1$) and $P_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{a_r+b_r} = 1$. We can summarize both scenarios as

$$P_N^{(\infty)}(\bar{\mathbf{w}}) = \begin{cases} N, & \text{if } \bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}, \quad \forall i \in \{1, \ldots, N\}, \\ 1, & \text{if } \bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}, \quad \forall i \in \{1, \ldots, N\}. \end{cases} \tag{63}$$

A.2 Special cases of $D_N^{(r)}(\bar{\mathbf{w}})$

Below, we analyze some special cases of the family $D_N^{(r)}(\bar{\mathbf{w}})$ in Eq. (42):

Case r→ 0. The coefficients of this family given in Table 2 reach the values $a_r \to a_0 = 0$ and $b_r \to b_0 = 1$. In this case, If $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}$, we have $\lim_{r\to 0} f_r(\bar{\mathbf{w}}) = 1$ (considering again $0^0 = 0$ and $1^\infty = 1$). Whereas, when $\bar{\mathbf{w}}$ is not a vertex, i.e., $\bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}$, then

$$\lim_{r\to 0} f_r(\bar{\mathbf{w}}) = \lim_{r\to 0} \left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{\frac{1}{r}} = \infty.$$

so that $\lim_{r\to 0} a_r f_r(\bar{\mathbf{w}})$ has the indeterminate form of type $0 \times \infty$ that can be converted to $\frac{\infty}{\infty}$ as shown below. We can write

$$\lim_{r\to 0} a_r \left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{\frac{1}{r}} = (1-N) \lim_{r\to 0} \frac{\left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{1/r}}{N^{1/r} - N}. \tag{64}$$

Moreover, when $r \to 0$ we have

$$\frac{\left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{1/r}}{N^{1/r} - N} \approx \frac{\left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{1/r}}{N^{1/r}} = \left[\frac{1}{N}\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{1/r}, \quad \text{when} \quad r \to 0. \tag{65}$$

For $r \to 0$, we can also write

$$(\bar{w}_n)^r = \exp(r \log \bar{w}_n) \approx 1 + r \log \bar{w}_n, \tag{66}$$

where we have used the Taylor expansion of first order of $\exp(r \log \bar{w}_n)$. Replacing $(\bar{w}_n)^r \approx 1 + r \log \bar{w}_n$ inside $\frac{1}{N}\sum_{n=1}^{N}(\bar{w}_n)^r$ we obtain

$$\frac{1}{N}\sum_{n=1}^{N}(\bar{w}_n)^r \approx \frac{1}{N}N + r\frac{1}{N}\sum_{n=1}^{N}\log \bar{w}_n = 1 + r\frac{1}{N}\log\prod_{n=1}^{N}\bar{w}_n \tag{67}$$

$$= 1 + r\log\left[\prod_{n=1}^{N}\bar{w}_n\right]^{\frac{1}{N}}. \tag{68}$$

Thus, we can write

$$\left[\frac{1}{N}\sum_{n=1}^{N}(\bar{w}_n)^r\right]^{1/r} \approx \left[1 + r\log\left[\prod_{n=1}^{N}\bar{w}_n\right]^{\frac{1}{N}}\right]^{1/r}. \tag{69}$$

Moreover, given $x \in \mathbb{R}$, for $r \to 0$ we have also the relationship

$$(1+rx)^{\frac{1}{r}} \to \exp(x),$$

by definition of exponential function. Replacing above, for $r \to 0$,

$$\left[1 + r\log\left[\prod_{n=1}^{N}\bar{w}_n\right]^{\frac{1}{N}}\right]^{1/r} \longrightarrow \exp\left(\log\left[\prod_{n=1}^{N}\bar{w}_n\right]^{\frac{1}{N}}\right) = \left[\prod_{n=1}^{N}\bar{w}_n\right]^{\frac{1}{N}}. \tag{70}$$

Thus, finally we obtain

$$\lim_{r\to 0} a_r \left[\sum_{n=1}^{N} (\bar{w}_n)^r\right]^{\frac{1}{r}} = (1-N)\left[\prod_{n=1}^{N}\bar{w}_n\right]^{1/N}, \tag{71}$$

and

$$\lim_{r\to 0} D_N^{(r)}(\bar{\mathbf{w}}) = \frac{1}{(1-N)\left[\prod_{n=1}^N \bar{w}_n\right]^{1/N}+1}, \tag{72}$$

$$D_N^{(0)}(\bar{\mathbf{w}}) = \frac{1}{(1-N)\mathrm{GeoM}(\bar{\mathbf{w}})+1} \tag{73}$$

Case r= 1. With a similar procedure used for $P_N^{(1)}$, we obtain $D_N^{(1)}(\bar{\mathbf{w}}) = P_N^{(1)}(\bar{\mathbf{w}})$.

Case r→ ∞. In this case, $a_r \to a_\infty = 1$ and $b_r \to b_\infty = 0$ and, since the distance $\left[\sum_{n=1}^N (\bar{w}_n)^r\right]^{\frac{1}{r}}$ converges to the L_∞ distance, $\max[\bar{w}_1, \ldots, \bar{w}_N]$, when $r \to \infty$ [22], we obtain $D_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{\max[\bar{w}_1,\ldots,\bar{w}_N]}$.

A.3 Special cases of $V_N^{(r)}(\bar{\mathbf{w}})$

In the following, we study some special cases of the family $V_N^{(r)}(\bar{\mathbf{w}})$ in Eq. (43):

Case r→ 0. The coefficients of this family given in Table 2 are $a_0 = 1$ and $b_0 = 0$. If $\bar{\mathbf{w}}$ does not contain zeros then $f_r(\bar{\mathbf{w}}) = \sum_{n=1}^N (\bar{w}_n)^r = N$. Otherwise, assuming $0^0 = 0$, If $\bar{\mathbf{w}}$ contains N_Z zeros, we have $f_r(\bar{\mathbf{w}}) = \sum_{n=1}^N (\bar{w}_n)^r = N - N_Z$. Thus, in general, we have

$$V^{(0)}(\bar{\mathbf{w}}) = N - N_Z.$$

Case r→ 1. In this, the coefficients a_r and b_r diverge. We can consider the limit

$$\begin{aligned}
&\lim_{r\to 1}\left(\frac{N^{r-1}(N-1)}{1-N^{r-1}}\left[\sum_{n=1}^N \bar{w}_n^r\right] + \frac{N^r-1}{N^{r-1}-1}\right) = \\
&= (N-1)\lim_{r\to 1}\frac{\left[\sum_{n=1}^N \bar{w}_n^r\right]-1}{1-N^{r-1}},
\end{aligned}$$

where we have a indetermination of type $\frac{0}{0}$. Using the L'Hôpital's rule [16], i.e., deriving both numerator and denominator w.r.t. r and computing the limit, we obtain (since $\log x = \frac{\log_2 x}{\log_2 e}$)

$$\begin{aligned}
\lim_{r\to 1}\frac{\left[\sum_{n=1}^N \bar{w}_n^r\right]-1}{1-N^{r-1}} &= \lim_{r\to 1}\frac{\left[\sum_{n=1}^N(\bar{w}_n^r \log \bar{w}_n)\right]}{-N^{r-1}\log N}, \\
&= \frac{\left[-\sum_{n=1}^N(\bar{w}_n \log \bar{w}_n)\right]}{\log N} = \frac{\left[-\sum_{n=1}^N(\bar{w}_n \log_2 \bar{w}_n)\right]}{(\log_2 e)\frac{\log_2 N}{\log_2 e}}, \\
&= \frac{H(\bar{\mathbf{w}})}{\log_2 N},
\end{aligned}$$

hence, finally,

$$V_N^{(1)}(\bar{\mathbf{w}}) = (N-1)\frac{H(\bar{\mathbf{w}})}{\log_2 N} + 1. \tag{74}$$

Case r→ ∞. The coefficients converge to the values $a_r \to a_\infty = 1 - N$ and $b_r \to b_\infty = N$. If $\bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}$ then $f_r(\bar{\mathbf{w}}) = \sum_{n=1}^{N} (\bar{w}_n)^r = 0$, so that $V_N^{(\infty)}(\bar{\mathbf{w}}) = b_\infty = N$. Otherwise, If $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}$, since $0^\infty = 0$ and considering $1^\infty = 1$, then $f_r(\bar{\mathbf{w}}) = \sum_{n=1}^{N} (\bar{w}_n)^r = 1$, so that $V_N^{(\infty)}(\bar{\mathbf{w}}) = a_\infty + b_\infty = 1$.

A.4 Special cases of $S_N^{(r)}(\bar{\mathbf{w}})$

Let us consider the family $S_N^{(r)}(\bar{\mathbf{w}})$. Four interesting special cases are studied below:

Case r→ 0. The coefficients given in Table 2 in this case are $a_r \to a_0 = 0$ and $b_r \to b_0 = 1$. If $\bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}$ then $f_r(\bar{\mathbf{w}}) \to \infty$, Otherwise, If $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}$, since $0^\infty = 0$ and considering $1^\infty = 1$, then $f_r(\bar{\mathbf{w}}) \to 1$. With a procedure similar to $D_N^{(0)}$, it is possible to show that

$$S^{(0)}(\bar{\mathbf{w}}) = (N^2 - N)\text{GeoM}(\bar{\mathbf{w}}) + 1. \tag{75}$$

Case r→ $\frac{1}{2}$. We have $a_{1/2} = 1$ and $b_{1/2} = 0$. Then, in this case, $S_N^{(\frac{1}{2})}(\bar{\mathbf{w}}) = f_{1/2}(\bar{\mathbf{w}}) = \left(\sum_{n=1}^{N} \sqrt{\bar{w}_n}\right)^2$.

Case r→ 1. With a similar procedure used for $V_N^{(1)}$, it is possible to obtain

$$S_N^{(1)}(\bar{\mathbf{w}}) = (N-1)\frac{H(\bar{\mathbf{w}})}{\log_2 N} + 1. \tag{76}$$

Case r→ ∞. In this case, $a_r \to a_\infty = -N$, $b_r \to b_\infty = N+1$. Moreover, $f_r(\bar{\mathbf{w}}) \to \max[\bar{w}_1, \dots, \bar{w}_n]$ [22], so that

$$S_N^{(\infty)}(\bar{\mathbf{w}}) = (N+1) - N\max[\bar{w}_1, \dots, \bar{w}_n]. \tag{77}$$

Table 5 summarizes all the special cases.

B Derivation of $Q_N(\bar{\mathbf{w}})$

Here, we derive the G-ESS function $Q_N(\bar{\mathbf{w}})$ given in Eq. (19) and induced by the L_1 distance. Let us define two disjoint sets of weights

$$\{\bar{w}_1^+, \dots, \bar{w}_{N^+}^+\} = \left\{\text{all } \bar{w}_n\text{:} \quad \bar{w}_n \geq 1/N, \quad \forall n = 1, \dots, N\right\}, \tag{78}$$
$$\{\bar{w}_1^-, \dots, \bar{w}_{N^-}^-\} = \left\{\text{all } \bar{w}_n\text{:} \quad \bar{w}_n < 1/N, \quad \forall n = 1, \dots, N\right\}, \tag{79}$$

Table 5: Special cases of the families $P_N^{(r)}(\bar{\mathbf{w}})$, $S_N^{(r)}(\bar{\mathbf{w}})$, $D_N^{(r)}(\bar{\mathbf{w}})$ and $V_N^{(r)}(\bar{\mathbf{w}})$.

Parameter:	$\mathbf{r \to 0}$	$\mathbf{r \to 1}$	$\mathbf{r = 2}$	$\mathbf{r \to \infty}$
$P_N^{(r)}(\bar{\mathbf{w}})$	$\frac{N}{N_Z+1}$ *Degenerate (type-1)*	$\frac{-N\log_2(N)}{-N\log_2(N)+(N-1)H(\bar{\mathbf{w}})}$ *Proper*	$\frac{1}{\sum_{n=1}^{N}(\bar{w}_n)^2}$ *Proper-Stable*	$\begin{cases} N, & \text{if } \bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}, \\ 1, & \text{if } \bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}. \end{cases}$ *Degenerate (type-1)*

Parameter:	$\mathbf{r \to 0}$	$\mathbf{r = \frac{1}{2}}$	$\mathbf{r \to 1}$	$\mathbf{r \to \infty}$
$S_N^{(r)}(\bar{\mathbf{w}})$	$(N^2-N)\mathrm{GeoM}(\bar{\mathbf{w}})+1$ *Degenerate (type-2)*	$\left(\sum_{n=1}^{N}\sqrt{\bar{w}_n}\right)^2$ *Proper-Stable*	$\frac{N-1}{\log_2(N)}H(\bar{\mathbf{w}})+1$ *Proper*	$N+1-N\max[\bar{w}_1,\ldots,\bar{w}_N]$ *Proper*

Parameter:	$\mathbf{r \to 0}$	$\mathbf{r \to 1}$	$\mathbf{r \to \infty}$
$D_N^{(r)}(\bar{\mathbf{w}})$	$\frac{1}{(1-N)\mathrm{GeoM}(\bar{\mathbf{w}})+1}$ *Degenerate (type-2)*	$\frac{-N\log_2(N)}{-N\log_2(N)+(N-1)H(\bar{\mathbf{w}})}$ *Proper*	$\frac{1}{\max[\bar{w}_1,\ldots,\bar{w}_N]}$ *Proper-Stable*

Parameter:	$\mathbf{r \to 0}$	$\mathbf{r \to 1}$	$\mathbf{r \to \infty}$
$V_N^{(r)}(\bar{\mathbf{w}})$	$N-N_Z$ *Degenerate (type-1)-Stable*	$\frac{N-1}{\log_2(N)}H(\bar{\mathbf{w}})+1$ *Proper*	$\begin{cases} N & \text{if } \bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}, \\ 1, & \text{if } \bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}. \end{cases}$ *Degenerate (type-1)*

where $N^+ = \#\{\bar{w}_1^+,\ldots,\bar{w}_{N^+}^+\}$ and $N^- = \#\{\bar{w}_1^-,\ldots,\bar{w}_{N^+}^-\}$. Clearly, $N^- + N^+ = N$ and $\sum_{i=1}^{N^+}\bar{w}_i^+ + \sum_{i=1}^{N^-}\bar{w}_i^- = 1$. Thus, we can write

$$\begin{aligned} L_1 &= \sum_{n=1}^{N}\left|\bar{w}_n - \frac{1}{N}\right| \\ &= \sum_{i=1}^{N^+}\left(\bar{w}_i^+ - \frac{1}{N}\right) + \sum_{j=1}^{N^-}\left(\frac{1}{N} - \bar{w}_j^-\right) \\ &= \sum_{i=1}^{N^+}\bar{w}_i^+ - \sum_{i=1}^{N^-}\bar{w}_i^- - \frac{N^+ - N^-}{N} \end{aligned} \tag{80}$$

and replacing the relationships $\sum_{i=1}^{N^+}\bar{w}_i^+ = 1 - \sum_{i=1}^{N^-}\bar{w}_i^-$ and $N^- = N - N^+$,

$$\begin{aligned} L_1 &= 2\left[\sum_{i=1}^{N^+}\bar{w}_i^+ - \frac{N^+}{N}\right], \\ &= 2\frac{N\sum_{i=1}^{N^+}\bar{w}_i^+ - N^+}{N}, \\ &= 2\left[\frac{Q_N(\bar{\mathbf{w}}) - N}{N}\right] + 2, \end{aligned} \tag{81}$$

where $Q_N(\bar{\mathbf{w}}) = -N\sum_{i=1}^{N^+}\bar{w}_i^+ + N^+ + N$.

C Relationships for $N = 2$

Other interesting relationships can be found for $N = 2$, i.e., when we have $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2]$ with $\bar{w}_2 = 1 - \bar{w}_1$. In this case, we can easily write two proper G-ESS functions, a quadratic (parabolic) function

$$\begin{aligned} \text{Par}(\bar{\mathbf{w}}) &= \text{Par}(\bar{w}_1), \\ &= 1 + 4\bar{w}_1 - 4\bar{w}_1^2, \qquad \bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2] \in \mathcal{S}_1, \end{aligned} \tag{82}$$

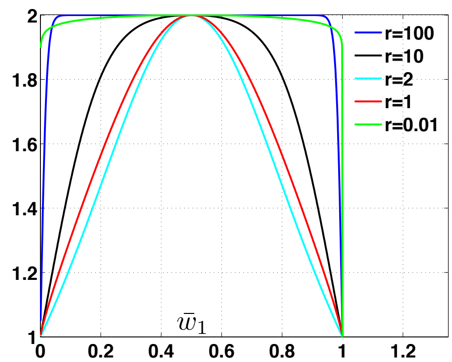
and a piecewise linear (triangular) function

$$\text{Tri}(\bar{\mathbf{w}}) = \text{Tri}(\bar{w}_1) = \begin{cases} 2\bar{w}_1 + 1, & \text{for } \bar{w}_1 \leq \frac{1}{2}, \\ 3 - 2\bar{w}_1, & \text{for } \bar{w}_1 > \frac{1}{2}, \end{cases} \qquad \bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2] \in \mathcal{S}_1. \tag{83}$$

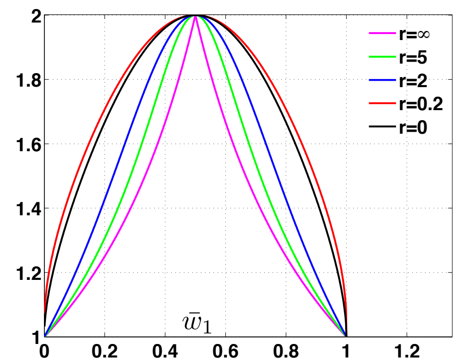
These two special cases are interesting since, for $N = 2$, there are several equivalences among different G-ESS functions, shown in Table 6. Some relationship in Table 6 can seem even surprising, but all the equivalences can be easily proved taking in account that $\bar{w}_2 = 1 - \bar{w}_1$. For instance, the standard formula $P_2^{(2)}$ for $N = 2$ is identical to the G-ESS function in Eq. (30) involving the harmonic mean. The second expression in Eq. (31) involving the harmonic mean coincides with the parabolic function $\text{Par}(\bar{\mathbf{w}})$ in Eq. (82). The G-ESS formula $S_2^{(\frac{1}{2})}(\bar{\mathbf{w}})$ for $N = 2$ is equivalent to $S_2^{(0)}(\bar{\mathbf{w}})$ involving the geometric mean.

Table 6: Equivalences among G-ESS functions for $N = 2$, $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2] \in \mathcal{S}_2$, i.e., with $\bar{w}_2 = 1 - \bar{w}_1$.

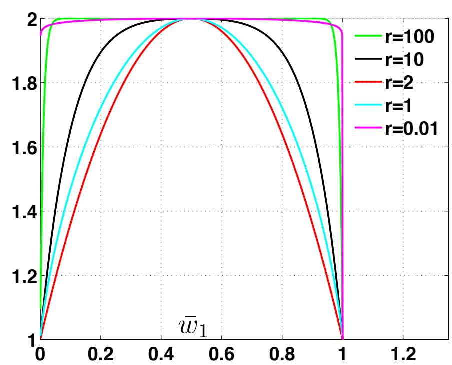
Equiv.	$P_2^{(2)}(\bar{\mathbf{w}}) \equiv A_{1,2}(\bar{\mathbf{w}})$	$A_{2,2}(\bar{\mathbf{w}}) \equiv \text{Par}(\bar{\mathbf{w}})$	$S_2^{(\frac{1}{2})}(\bar{\mathbf{w}}) \equiv S_2^{(0)}(\bar{\mathbf{w}})$	$S_2^{(\infty)}(\bar{\mathbf{w}}) \equiv T_{2,2}(\bar{\mathbf{w}}) \equiv \text{Tri}(\bar{\mathbf{w}})$	$D_2^{(\infty)}(\bar{\mathbf{w}}) \equiv T_{1,2}(\bar{\mathbf{w}})$
Ref.	Eq. (6); Eq. (30)	Eq. (31); Eq. (82)	Table 5	Table 5; Eq. (33); Eq. (83)	Table 5; Eq. (32)



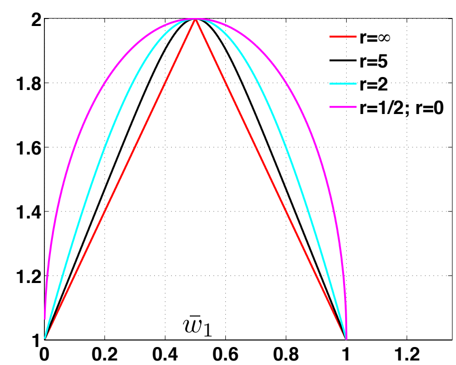
(a) $P_N^{(r)}$ with $N = 2$.



(b) $D_N^{(r)}$ with $N = 2$.

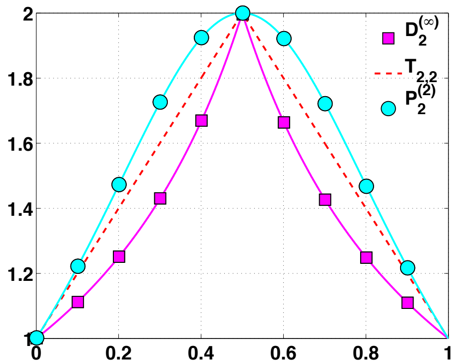


(c) $V_N^{(r)}$ with $N = 2$.

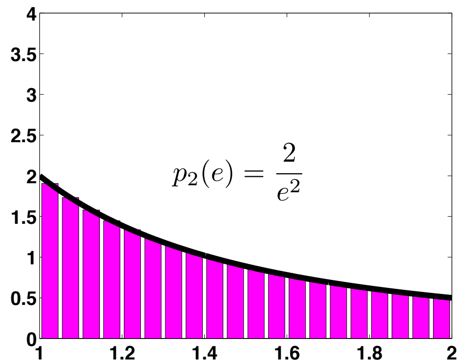


(d) $S_N^{(r)}$ with $N = 2$.

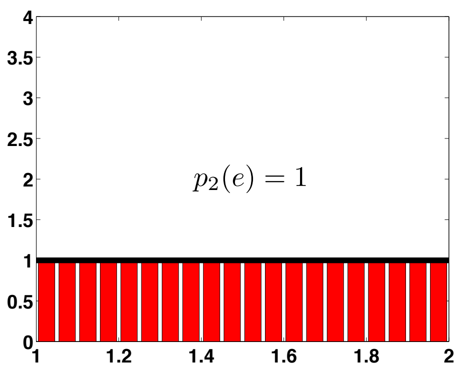
Figure 5: Different G-ESS functions as functions of $\bar{w}_1$, with $N = 2$ and different values of r.



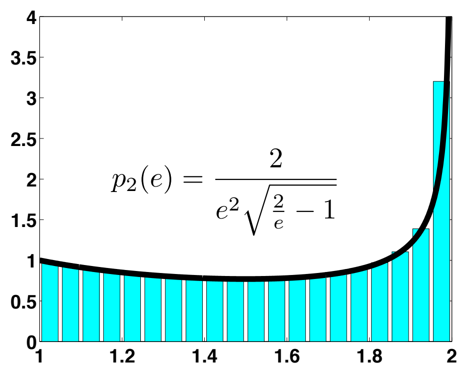
(a) $D_2^{(\infty)}$, $P_2^{(2)}$ and $T_{2,2}$ versus $\bar{w}_1$.



(b) Pdf $p_N(e)$ associated to $D_2^{(\infty)}$.

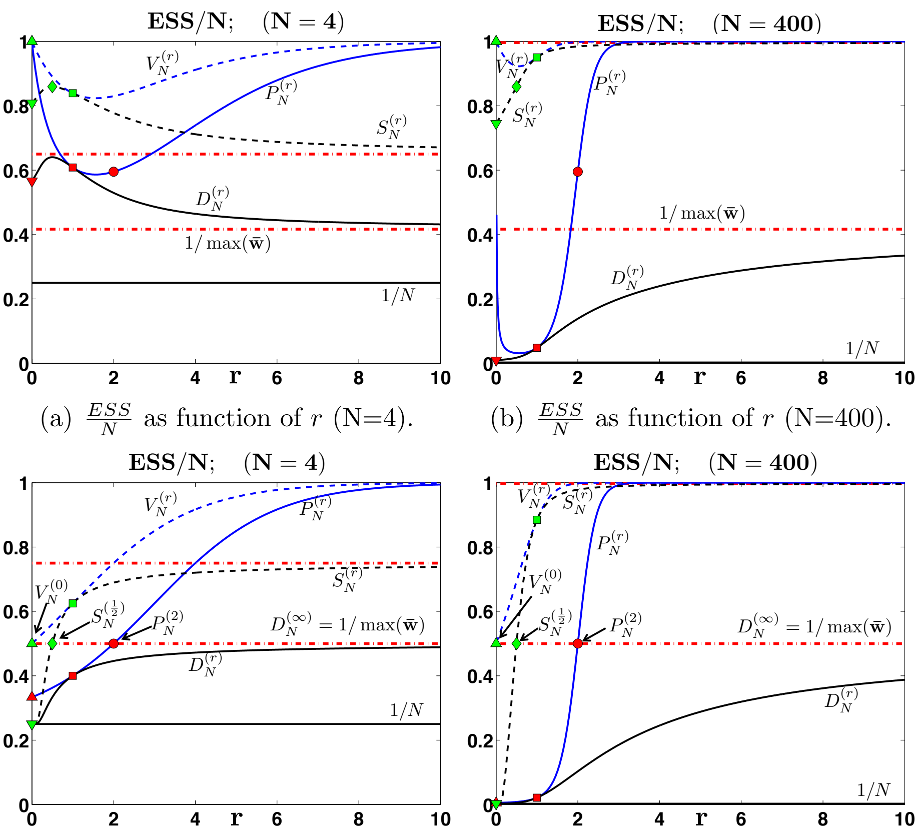


(c) Pdf $p_N(e)$ associated to $T_{2,2}$.



(d) Pdfs $p_N(e)$ associated to $P_2^{(2)}$.

Figure 6: **(a)** G-ESS functions, $D_2^{(\infty)}$ in Eq. (17) (squares), $P_2^{(2)}$ in Eq. (6) (circles), and $T_{2,2}$ in Eq. (33) (dashed lines), all of them with $N = 2$ (then, $\bar{w}_2 = 1 - \bar{w}_1$). We can see that $D_2^{(\infty)}$ has a sub-linear increase to the value $N = 2$, whereas $P_2^{(2)}$ a super-linear increase. **(b)-(c)-(d)** Pdfs $p_N(e)$ associated to $D_2^{(\infty)}$, $P_2^{(2)}$ and $T_{2,2}$, respectively. For $D_2^{(\infty)}$ (Fig. (b)) more probability mass is located close to 1, whereas for $P_2^{(2)}$ (Fig. (d)), $p_2(e)$ is unbalanced to the right side close to 2.



(a) $\frac{ESS}{N}$ as function of r (N=4). (b) $\frac{ESS}{N}$ as function of r (N=400).

(c) $\frac{ESS}{N}$ as function of r (N=4). (d) $\frac{ESS}{N}$ as function of r (N=400).

Figure 7: The rate $\frac{1}{N} \leq \frac{ESS}{N} \leq 1$ as function of r for the different families $P_N^{(r)}$, $D_N^{(r)}$ (both in solid lines), $V_N^{(r)}$ and $S_N^{(r)}$ (both in dashed lines), considering the vectors $\bar{\mathbf{w}} = [0.1\ 0.1\ 0.2\ 0.6]$ in **(a)**, its repeated version (100 times) in **(b)**, the $\bar{\mathbf{w}} = [0\ 0\ 0.5\ 0.5]$ in **(c)** and its repeated version (100 times) in **(d)**. The circle corresponds to $P_N^{(2)}$, the rhombus represents $S_N^{(\frac{1}{2})}$, $D_N^{(\infty)}$ is shown with a dotted straight line. The G-ESS functions involving the discrete entropy $H(\bar{\mathbf{w}})$ (i.e., $r \to 1$) and geometric mean $\text{GeoM}(\bar{\mathbf{w}})$ are depicted with squares and triangles down, respectively. The other two cases for $r \to 0$ are shown with triangles up (in this case, both coincide to 1).