Effective sample size for importance sampling based on discrepancy measures

Luca Martinoa, Víctor Elvirab, Francisco Louzadaa

aUniversidade de São Paulo, São Carlos (Brazil)
bUniversidad Carlos III de Madrid, Leganés (Spain)

1. Introduction

Sequential Monte Carlo (SMC) methods (a.k.a., particle filtering algorithms) are important tools for Bayesian inference [10], extensively applied in signal processing [9,17,29,26] and statistics [11,30,35]. 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 [9,11]. However, the application of resampling yields loss of diversity in the population of particles and entails an additional computational cost [9,12,6, p. 21]. Furthermore, resampling limits the parallel implementation of the filter (since it needs the information of all the weights at a specific iteration). Hence, one would desire to perform resampling steps parsimoniously, only when it is strictly required [12, pp. 13 and 15]. This adaptive implementation of the resampling procedure needs the use of the concept of Effective Sample Size (ESS) of a set of weighted samples [9,23,35].

The ESS is a measure of the efficiency of different Monte Carlo methods, such as Markov Chain Monte Carlo (MCMC) and Importance Sampling (IS) techniques [4,15,23,35,25,27]. ESS is theoretically defined as the equivalent number of independent samples generated directly from the target distribution, which yields the same efficiency in the estimation obtained by the MCMC or IS algorithms. Thus, a possible mathematical definition [15,21] considers the ESS as a 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 the same number of samples in both estimators.

The most common choice to approximate this theoretical ESS definition in IS is the formula

\[ \text{ESS} = \frac{1}{\sum_{n=1}^{N} w_n^2} \]

This expression is obtained by several approximations of the initial theoretical definition so that, ESS provides an accurate estimation of the ESS values (given by the theoretical definition) only in specific cases. For this reason other ESS expressions have also been proposed, e.g., the perplexity, involving the discrete entropy of the weights [7] has been suggested in [2]; see also [35, Chapter 4], [12, Section 3.5], [18]. The discrete entropy has also been considered in order to design criteria for adaptive resampling schemes in [32, Section 2.3], [31]. More recently, other alternative formulas ESS have also been analyzed in [18]. In [37], a conditional ESS formula is introduced in order to study similarities between successive pdf’s within a sequence of densities.

However, the ESS approximation \[ \text{ESS} = \frac{1}{\sum_{n=1}^{N} w_n^2} \] is widely used in...
practice, and it generally provides good performance. Furthermore, several theoretical studies related to \( \hat{ESS} \) can be found in literature (e.g., see [1,36,32,31]). It is possible to show that \( \hat{ESS} \) is also related to the discrepancy between the probability mass function (pmf) defined by the normalized weights \( \hat{w}_n \), \( n = 1, \ldots, N \), and the uniform pmf \( \mathcal{U}(1, 2, \ldots, N) \). When the pmf defined by \( \hat{w}_n \) is close to the uniform pmf \( \mathcal{U}(1, 2, \ldots, N) \), \( \hat{ESS} \) provides high values otherwise, when the pmf defined by \( \hat{w}_n \) is concentrated mainly in one weight, \( \hat{ESS} \) provides small values. More specifically, we show that \( \hat{ESS} \) is related to the Euclidean distance between these two pmfs.

It is possible to obtain other ESS functions based on different discrepancy measures, as we show in this work. 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 the geometric mean, discrete entropy [7] and the Gini coefficient [16,24] of the normalized weights, are presented. Additionally, four families of proper G-ESS functions are designed. We classify the novel G-ESS functions (including also the perplexity measure [2,35]) according to the conditions fulfilled. We focus on the G-ESS functions which satisfy all the desirable conditions and compare them by means of numerical simulations. This analysis shows that different G-ESS expressions present interesting features from a theoretical and practical point of view and it can be considered valid alternatives of the standard formula

\[
\frac{1}{\sum_{i=1}^{N} \hat{w}_i}.
\]

The rest of the paper is organized as follows. Section 2 recalls the required background material. In Section 3, we highlight that the standard formula \( \hat{ESS} = \frac{1}{\sum_{i=1}^{N} \hat{w}_i} \) is related to the euclidean distance between two pmfs. The definition of a generalized ess function is given in Section 4, and novel ess families are introduced in Section 5. Section 6 provides several numerical simulations. Finally, Section 7 contains some brief conclusions.

2. Effective sample size for importance sampling

Let us denote the target pdf as \( \pi(x) \propto p(x) \) (known up to a normalizing constant) with \( x \in X \). Moreover, we consider the following integral involving \( \pi(x) \) and a square-integrable (w.r.t. \( \pi \)) function \( h(x) \),

\[
l = \int_X h(x) \pi(x) \, dx,
\]

that we desire to approximate using a Monte Carlo approach. If we are able to draw \( N \) independent samples \( x_1, \ldots, x_N \) from \( \pi(x) \), then the Monte Carlo estimator of \( l \) is

\[
\hat{l} = \frac{1}{N} \sum_{n=1}^{N} h(x_n) \approx l,
\]

where \( x_n \sim \pi(x) \), with \( n = 1, \ldots, N \). However, in general, generating samples directly from the target, \( \pi(x) \), is impossible. Alternatively, we can draw \( N \) samples \( x_1, \ldots, x_N \) from a simpler proposal pdf \( q(x) \),\(^2\) and then assign a weight \( w_n = \frac{p(x_n)}{q(x_n)} \), \( n = 1, \ldots, N \), to each sample, according to the importance sampling (IS) approach. Defining the normalized weights,

\[
\hat{w}_n = \frac{w_n}{\sum_{i=1}^{N} w_i}, \quad n = 1, \ldots, N,
\]

then the IS estimator is

\[
\hat{l} = \frac{N}{\sum_{n=1}^{N} \hat{w}_n} h(x_n) \approx l,
\]

with \( x_n \sim q(x) \), \( n = 1, \ldots, N \). In general, the estimator \( \hat{l} \) is less efficient than \( l \), since the samples are not directly generated by \( \pi(x) \). In several applications [9,11,17,26], it is necessary to measure in some way the efficiency that we lose using \( \hat{l} \) instead of \( l \). The idea is to define the Effective Sample Size (ESS) as ratio of the variances of the estimators [21],

\[
ESS = \frac{N \text{var}[\hat{l}]}{\text{var}[l]}. \tag{5}
\]

Remark. The ESS value in (5) can be interpreted as the number of independent samples drawn directly from the target \( \pi \) required in order to obtain an estimator \( \hat{l} \) with a variance equal to var[\( l \)].

Namely, ESS represents the number of samples from \( \pi \) required to obtain a Monte Carlo estimator \( \hat{l} \) with the same efficiency of the IS estimator \( l \) (considering \( q(x) \) as proposal). Heuristically speaking, we can assert that ESS measures how many independent identically distributed (i.i.d.) samples, drawn from \( \pi \), are equivalent to the \( N \) weighted samples, drawn from \( q(x) \) and weighted according to the ratio \( \frac{\pi(x)}{q(x)} \) [5, Section 3].

Finding a useful expression of ESS derived analytically from the theoretical definition above is not straightforward. Different derivations, [21, 22, 11, Chapter 11; 35, Chapter 4] proceed using several approximations and assumptions for yielding an expression useful from a practical point of view. A well-known ESS approximation, widely used in the literature [11,23,35], is

\[
\hat{ESS} = \frac{1}{\sum_{i=1}^{N} \hat{w}_i^2} \approx \frac{\sum_{i=1}^{N} w_i^2}{\sum_{i=1}^{N} \hat{w}_i^2}, \quad ESS \equiv P_{aw}^2(w), \tag{6}
\]

where we have used the normalized weights \( w = [\hat{w}_1, \ldots, \hat{w}_N] \) in the first equality, and the unnormalized ones in the second equality. The reason of using the notation \( P_{aw}^2(w) \) 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 5). An interesting property of the expression (6) is that

\[
1 \leq P_{aw}^2(w) \leq N. \tag{7}
\]

3. \( P_{aw}^2(w) \) as a discrepancy measure

Although, in the literature, \( P_{aw}^2(w) \) is often considered a suitable approximation of the theoretical ESS definition, the derivation of \( P_{aw}^2 \) [21,34,35,5, Section 3] contains several approximations and strong assumptions. As a consequence, \( P_{aw}^2 \) differs substantially from the original definition \( ESS = N \frac{\text{var}[l]}{\text{var}[\hat{l}]} \) in many scenarios (e.g., see the numerical results in Section 6.2). In Appendix A, we list the approximations needed in the derivation of \( P_{aw}^2 \) and we also discuss its limitations.

Despite the previous consideration, the expression \( P_{aw}^2(w) \) is widely used in the adaptive resampling context [9,12, 6, p. 21] within population Monte Carlo and particle filtering schemes [3,11,14,17,26].\(^3\) For this reason, several theoretical studies about

\(^2\) We assume that \( p(x) > 0 \) for all \( x \) where \( \pi(x) \neq 0 \), and \( \pi(x) \) has heavier tails than \( p(x) \).

\(^3\) In a standard resampling procedure [12,6], the indices of the particles employed at the next generation are drawn according to a multinomial distribution defined by the normalized weights \( \hat{w}_n = \frac{w_n}{\sum_{i=1}^{N} w_i} \), \( n = 1, \ldots, N \). In order to perform resampling steps adaptively, i.e., only in certain specific iterations, the
can also be found in literature [1,36,32,31], showing that \( P_N^{(2)} \) has good theoretical properties (e.g., monitoring \( P_N^{(2)} \) is enough to prevent the particle system to collapse [1,31,32]).

We believe that one of the reasons of the success of \( P_N^{(2)} \) in adaptive resampling is due to its connection with the discrepancy between two pmfs: the pmf defined by the weights \( \mathbf{w} = [w_1, ..., w_N] \) and the discrete uniform pmf defined by \( \mathbf{w}^* = \left[ \frac{1}{N}, ..., \frac{1}{N} \right] \). Roughly speaking, if the vector \( \mathbf{w} \) is reasonably close to \( \mathbf{w}^* \), then the resampling is considered unnecessary. Otherwise, the resampling is applied. More precisely, we can show that \( P_N^{(2)} \) is related to the Euclidean distance \( L_2 \) between these two pmfs, i.e.,

$$
\| \mathbf{w} - \mathbf{w}^* \|_2 = \left( \sum_{n=1}^{N} \left( w_n - \frac{1}{N} \right)^2 \right)^{1/2}
$$

Hence, maximizing \( P_N^{(2)} \) is equivalent to minimizing the Euclidean distance \( \| \mathbf{w} - \mathbf{w}^* \|_2 \). Thus, it appears natural to consider the possibility of using other discrepancy measures between these pmfs, in order to derive alternative ESS functions. In Appendix C, we show other possible ESS expressions induced by non-Euclidean distances. In the following, we define a generic ESS function through the introduction of five conditions (three of them strictly required, and two welcome conditions), and then we provide several examples.

4. Generalized ESS functions

In this section, we introduce some properties that a Generalized ESS (G-ESS) function, based only on the information of the normalized weights, should satisfy. Here, first of all, note that any possible G-ESS is a function of the vector of normalized weights \( \mathbf{w} = [w_1, ..., w_N] \),

$$
E_N(\mathbf{w}) = E_N(w_1, ..., w_N): S_N \rightarrow [1, N],
$$

where \( S_N \subset \mathbb{R}^N \) represents the unit simplex in \( \mathbb{R}^N \). Namely, the variables \( w_1, ..., w_N \) are subjected to the constraint

$$
\sum_{n=1}^{N} w_n = 1.
$$

4.1. Conditions for the G-ESS functions

Below we list five conditions that \( E_N(\mathbf{w}) \) should fulfill to be considered 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.,

$$
E_N(w_1, w_2, ..., w_N) = E_N(w_{j_1}, w_{j_2}, ..., w_{j_N}),
$$

for any possible set of indices \( (j_1, ..., j_N) = (1, ..., N) \).

C2. Maximum condition: A maximum value is \( N \) and it is reached at \( \mathbf{w}^* = \left[ \frac{1}{N}, ..., \frac{1}{N} \right] \) (see Eq. (B.1)), i.e.,

$$
E_N(\mathbf{w}^*) = N \geq E_N(\mathbf{w}),
$$

C3. Minimum condition: the minimum value is 1 and it is reached (at least) at the vertices \( \mathbf{w}^{(j)} = \left[ 0, 0, ..., 1, 0, ..., 0 \right] \) of the unit simplex in Eq. (B.2),

$$
E_N(\mathbf{w}^{(j)}) = 1 \leq E_N(\mathbf{w}),
$$

for all \( j \in \{1, ..., N\} \).

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

C5. Stability—Invariance of the rate \( \frac{E_N(\mathbf{w})}{E_N(\mathbf{w}^*)} \): A maximum value is 1 and it is reached at \( \mathbf{w}^* \). The invariance condition is expressed as

$$
\frac{E_N(\mathbf{w})}{N} = \frac{E_{MN}(\mathbf{w})}{MN} = \frac{E_N(\mathbf{w})}{M} = \frac{E_N(\mathbf{w})}{M},
$$

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

The condition C5 is related to the optimistic approach described in Appendix B. For clarifying this point, as an example, let us consider the vectors

$$
\mathbf{w} = [0, 1, 0], \quad \mathbf{v} = \left[ 0, \frac{1}{2}, 0, 0, \frac{1}{2}, 0 \right] = \frac{1}{2} [\mathbf{w}, \mathbf{w}],
$$

$$
\mathbf{v}^* = \left[ 0, \frac{1}{2}, 0, 0, \frac{1}{2}, 0 \right] = \frac{1}{2} [\mathbf{w}, \mathbf{w}],
$$

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

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

4.2. Classification 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

<table>
<thead>
<tr>
<th>Class of G-ESS</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degenerate (D)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Proper (P)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Degenerate and Stable (DS)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Proper and Stable (PS)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(footnote continued)
Table 2
G-ESS families and their coefficients $a_i$ and $b_i$.

<table>
<thead>
<tr>
<th>$P_i^\theta(\mathbf{w})$</th>
<th>$D_i^\theta(\mathbf{w})$</th>
<th>$V_i^\theta(\mathbf{w})$</th>
<th>$S_i^\theta(\mathbf{w})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{a_i \sum_{n=1}^{N} (w_n)^i + b_i}$</td>
<td>$\frac{1}{a_i \sum_{n=1}^{N} (w_n)^i + b_i}$</td>
<td>$a_i \sum_{n=1}^{N} (w_n)^i + b_i$</td>
<td>$a_i \left( \sum_{n=1}^{N} (w_n)^i \right)^{-1} + b_i$</td>
</tr>
<tr>
<td>$a_i = \frac{1 - N}{N^{2-(r-1)} - N}$</td>
<td>$a_i = \frac{N - 1}{N - N^{r}}$</td>
<td>$a_i = \frac{N^{-r} - (N - 1)}{N^{-r} - 1}$</td>
<td>$a_i = \frac{N - 1}{N^{r} - 1}$</td>
</tr>
<tr>
<td>$b_i = \frac{N^{2-(r-1)} - 1}{N^{2-(r-1)} - N}$</td>
<td>$b_i = \frac{1 - N^{r}}{N - N^{r}}$</td>
<td>$b_i = \frac{N^{-r} - 1}{N^{-r} - 1}$</td>
<td>$b_i = \frac{N^{r} - N}{N^{r} - 1}$</td>
</tr>
</tbody>
</table>

Table 3
Special cases of the families $P_i^\theta(\mathbf{w})$, $S_i^\theta(\mathbf{w})$, $D_i^\theta(\mathbf{w})$ and $V_i^\theta(\mathbf{w})$.

<table>
<thead>
<tr>
<th>Parameter:</th>
<th>$r \to 0$</th>
<th>$r \to 1$</th>
<th>$r \to \infty$</th>
<th>$r \to 0$</th>
<th>$r = \frac{1}{2}$</th>
<th>$r \to 1$</th>
<th>$r \to \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_i^\theta(\mathbf{w})$</td>
<td>$\frac{N}{N^{-r}}$</td>
<td>$\frac{-N \log(N)}{1 - \frac{\log(N)}{N!}}$</td>
<td>$\frac{1}{\sum_{n=1}^{N} (w_n)^i}$</td>
<td>${ N, \text{ if } \mathbf{w} \neq \mathbf{w}^{(i)} }$</td>
<td>$1, \text{ if } \mathbf{w} = \mathbf{w}^{(i)} }$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degenerate (type-1)</td>
<td>Proper</td>
<td>Proper-Stable</td>
<td>Proper</td>
<td>Degenerate (type-1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_i^\theta(\mathbf{w})$</td>
<td>$(N^2 - N \log \text{Geo}(\mathbf{w}) + 1$</td>
<td>$(\sum_{n=1}^{N} \sqrt{w_n})^2$</td>
<td>$\frac{N - 1}{N \log(N)}$</td>
<td>$1, \text{ if } \mathbf{w} = \mathbf{w}^{(i)} }$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degenerate (type-2)</td>
<td>Proper-Stable</td>
<td>Proper</td>
<td>Proper</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_i^\theta(\mathbf{w})$</td>
<td>$\frac{1}{1 - N \log \text{Geo}(\mathbf{w}) + 1}$</td>
<td>$\frac{-N \log(N)}{1 - \frac{\log(N)}{N!}}$</td>
<td>$\frac{1}{\max(w_n, \ldots, w_N)}$</td>
<td>${ N, \text{ if } \mathbf{w} \neq \mathbf{w}^{(i)} }$</td>
<td>$1, \text{ if } \mathbf{w} = \mathbf{w}^{(i)} }$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degenerate (type-2)</td>
<td>Proper-Stable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_i^\theta(\mathbf{w})$</td>
<td>$N - N^2$</td>
<td>$\frac{N - 1}{\log(N)}$</td>
<td>$1, \text{ if } \mathbf{w} = \mathbf{w}^{(i)} }$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degenerate (type-1)-Stable</td>
<td>Proper</td>
<td>Degenerate (type-1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Stable G-ESS functions.

<table>
<thead>
<tr>
<th>$D_i^{\text{ess}}(\mathbf{w})$</th>
<th>$P_i^{\text{ess}}(\mathbf{w})$</th>
<th>$V_i^{\text{ess}}(\mathbf{w})$</th>
<th>$S_i^{\text{ess}}(\mathbf{w})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{\max(w_0, \ldots, w_N)}$</td>
<td>$\frac{1}{\sum_{n=1}^{N} \sqrt{w_n}}$</td>
<td>$(\sum_{n=1}^{N} \sqrt{w_n})^2$</td>
<td>$N - N^2$</td>
</tr>
<tr>
<td>proper</td>
<td>proper</td>
<td>proper</td>
<td>degenerate (type-1)</td>
</tr>
<tr>
<td>$Q_{\text{ess}}(\mathbf{w})$</td>
<td>$\frac{N - \text{plus}<em>{\log}(\mathbf{w})}{N - \text{plus}</em>{\log}(\mathbf{w})}$</td>
<td>$\frac{\log(N)}{1 - \frac{\log(N)}{N!}}$</td>
<td>$\frac{1}{\max(w_n, \ldots, w_N)}$</td>
</tr>
<tr>
<td>$-N \sum_{n=1}^{N} (w_n)^i + N^i + N$</td>
<td>$N^i - N \log(\mathbf{w}) + N^i$</td>
<td>$\frac{N^i}{2\log(\mathbf{w})}$</td>
<td>proper</td>
</tr>
<tr>
<td>proper</td>
<td>degenerate (type-2)</td>
<td>proper</td>
<td></td>
</tr>
</tbody>
</table>

G-ESS functions which satisfy the first three conditions, C1, C2 and C3 but not 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 $E_{\theta}(\mathbf{w})$ reaches the maximum value $N$ also in some other point $\mathbf{w} \neq \mathbf{w}^{(i)}$, or type-2 if $E_{\theta}(\mathbf{w})$ reaches the minimum value 1 also in some point that is not a vertex.

5. G-ESS families and further examples

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

$$P_i^\theta(\mathbf{w}) = \frac{1}{a_i \sum_{n=1}^{N} (w_n)^i + b_i}, \quad r \in \mathbb{R},$$

$$D_i^\theta(\mathbf{w}) = \frac{1}{a_i \left( \sum_{n=1}^{N} (w_n)^i \right)^{-1} + b_i}, \quad r \geq 0,$$

$$V_i^\theta(\mathbf{w}) = a_i \sum_{n=1}^{N} (w_n)^i + b_i, \quad r \in \mathbb{R},$$

$$S_i^\theta(\mathbf{w}) = a_i \left( \sum_{n=1}^{N} (w_n)^i \right)^{-1} + b_i, \quad r \geq 0,$$
where $a_r$ and $b_r$ are constant values depending on the parameter $r$ (and the corresponding family). The values of the coefficients $a_r$ and $b_r$ can be found easily as solutions of linear systems (see Appendix D), 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

\[
P_{N}^r(e) = \frac{N-1}{N^{1/2} - 1} \left(1 - \frac{\sum_{n=1}^{N} w_n^r}{N^{1/2}} \right) + 1 - \frac{N-1}{N^{1/2} - 1}.
\]

(20)

These families contain different G-ESS functions previously introduced, and also other interesting special cases. Table 3 summarizes these particular cases (jointly with the corresponding classification) corresponding to specific values the parameter $r$. Some of them ($D_{2}^{(\infty)}$ and $S_{2}^{(P)}$) involve the geometric mean of the normalized weights,

\[
\text{GeoM}(\mathbf{w}) = \left(\prod_{n=1}^{N} w_n\right)^{1/N}.
\]

(21)

other ones ($D_{2}^{(P)}$ and $S_{2}^{(V)}$) involve the discrete entropy [7] of the normalized weights,

\[
H(\mathbf{w}) = -\sum_{n=1}^{N} w_n \log_2(w_n).
\]

(22)

and others use the number of zeros contained in $\mathbf{w}$, $N_{0} = \#(w_n = 0, \forall n = 1, \ldots, N)$. The derivations of these special cases are provided in Appendices D.1 and D.2. Note that Table 3 contains a proper and stable G-ESS function
Other examples of G-ESS functions, which do not belong to these families, are given below.

**Example 1.** The following function

\[ Q_N(\mathbf{w}) = -N \sum_{i=1}^{N^*} w_i^+ + N^* + N, \]

with

\[ S_1^{1/2}(\mathbf{w}) = \left( \sum_{n=1}^{N} \sqrt{w_n} \right), \]

not introduced so far. Other examples of G-ESS functions, which do not belong to these families, are given below.

**Example 2.** The following functions involving the minimum of the normalized weights,

\[ T_{1,N}(\mathbf{w}) = \frac{1}{(1 - N) \min\{w_1^+, \ldots, w_N^+\} + 1}, \]

with

Table 5

<table>
<thead>
<tr>
<th>Description</th>
<th>N</th>
<th>( D_N^{(\infty)} / N )</th>
<th>( P_N^{(2)} / N )</th>
<th>( S_N^{(1/2)} / N )</th>
<th>( Q_N / N )</th>
<th>( \text{Gini}_N / N )</th>
<th>( \text{Per}_N / N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>50</td>
<td>0.2356</td>
<td>0.5194</td>
<td>0.7902</td>
<td>0.6371</td>
<td>0.5117</td>
<td>0.6555</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.1776</td>
<td>0.5057</td>
<td>0.7868</td>
<td>0.6326</td>
<td>0.5020</td>
<td>0.6568</td>
</tr>
<tr>
<td></td>
<td>( 10^3 )</td>
<td>0.1366</td>
<td>0.5013</td>
<td>0.7858</td>
<td>0.6324</td>
<td>0.5007</td>
<td>0.6558</td>
</tr>
<tr>
<td></td>
<td>5 \times 10^3</td>
<td>0.1121</td>
<td>0.5005</td>
<td>0.7856</td>
<td>0.6322</td>
<td>0.5002</td>
<td>0.6554</td>
</tr>
<tr>
<td>Std</td>
<td>50</td>
<td>0.0517</td>
<td>\textbf{0.0622}</td>
<td>0.0324</td>
<td>0.0345</td>
<td>0.0410</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0336</td>
<td>\textbf{0.0341}</td>
<td>0.0168</td>
<td>0.0171</td>
<td>0.0204</td>
<td>0.0248</td>
</tr>
<tr>
<td></td>
<td>( 10^3 )</td>
<td>\textbf{0.0213}</td>
<td>0.0158</td>
<td>0.0077</td>
<td>0.0077</td>
<td>0.0091</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>5 \times 10^3</td>
<td>\textbf{0.0145}</td>
<td>0.0071</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0040</td>
<td>0.0059</td>
</tr>
</tbody>
</table>

Fig. 2. The histograms of the rates \( \frac{Q_{\infty}}{N} \) corresponding to the proper and stable G-ESS functions in Table 4, with \( N \in \{50, 1000\} \).
Example 3. The perplexity function introduced in [2] and also contained in a ESS family studied in [18, pp. 13 and 22] is defined as

\[ \text{Per}_N(w) = 2^{H(w)} \]

where

\[ H(w) = - \sum_{n=1}^{N} w_n \log_2(w_n) \]

is the discrete entropy [7] of the pmf \( w_n \), \( n = 1, \ldots, N \). The perplexity is a proper and stable G-ESS function.

Example 4. Let us consider the Gini coefficient \( G(w) [16, 24] \), defined as follows. First of all, we define the non-decreasing sequence of normalized weights

\[ \bar{w}_1 \leq \bar{w}_2 \leq \cdots \leq \bar{w}_N \]

obtained by sorting in ascending order the entries of the vector \( w \).

The Gini coefficient is defined as

\[ G(w) = 2 \frac{s(w)}{N} - N + 1 \]

where

\[ s(w) = \sum_{n=1}^{N} m_{\bar{w}_n} \]

Then, the G-ESS function defined as

\[ \text{Gini}_N(w) = -NG(w) + N \]

is proper and stable.

Example 5. The following G-ESS function (inspired by the \( L_1 \) distance),

\[ N - \text{plus}_N(w) = N^+ = \# \{ \bar{w}_n \geq 1/N, \quad \forall \ n = 1, \ldots, N \} \]
is also degenerate (type 2) and stable.

5.1. 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 4. The following ordering inequalities

\[ P_0^{\infty}(\mathbf{w}) \leq P_0^{1}(\mathbf{w}) \leq V_0^{1}(W), \quad \forall \mathbf{w} \in S_N, \]

can also be proved.

5.2. 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 \( \mathbf{w} \) is considered as a realization of a random variable uniformly distributed in the unit simplex \( S_N \). Namely, let us consider the random variables \( \mathbf{W} \sim U(S_N) \) and \( E = E(t)(\mathbf{W}) \) with probability density function (pdf) \( p_N(e) \), i.e.,

\[ E \sim p_N(e). \]  

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 \( E_{\alpha} \) concentrate more probability mass closer to the maximum \( N \), other functions closer to the minimum 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 \( E_{\alpha}(W_1, 1 - W_1) = E(W_1) \), then we obtain

\[ p_2(e) = \frac{dI_1(e)}{de} + \frac{dI_2(e)}{de}, \quad e \in [1, N], \]  

using the expression of transformation of a uniform random variable, defined in \([0, 1]\). Thus, we find that \( p_2(e) = \frac{2}{e} \) for \( D_1^{\infty} \) and

\[ \begin{align*}
\text{(a) } & \mu_p = 0.3 \text{ (and } \sigma_p = 1) \text{.} \\
\text{(b) } & \mu_p = 0.5 \text{ (and } \sigma_p = 1) \text{.} \\
\text{(c) } & \mu_p = 1 \text{ (and } \sigma_p = 1) \text{.} \\
\text{(d) } & \mu_p = 1.5 \text{ (and } \sigma_p = 1) \text{.}
\end{align*} \]
for instance. Fig. 1 depicts the pdf’s $p_e$ for $D_2^{(\infty)}$, $T_{2,2}$ in Eq. (26) and $P_2^{(i)}$ in Eq. (6). We can observe that $P_2^{(i)}$ is more optimistic than $D_2^{(\infty)}$ judging a set of weighted samples and assigning a value of the effective size, since $p_e$ in this case is unbalanced to the right side close to 2. From a practical point of view, the pdf $p_{\bar{e}}$ could be used for choosing the threshold values for the adaptive resampling. The limiting distribution obtained for $N \to \infty$,

$$p_{\bar{e}}(e) = \frac{2}{e^{\sqrt{2}}-1}$$

for $P_2^{(i)}$, for instance. Fig. 1 depicts the pdf’s $p_{\bar{e}}$ for $D_2^{(\infty)}$, $T_{2,2}$ in Eq. (26) and $P_2^{(i)}$ in Eq. (6). We can observe that $P_2^{(i)}$ is more optimistic than $D_2^{(\infty)}$ judging a set of weighted samples and assigning a value of the effective size, since $p_e$ in this case is unbalanced to the right side close to 2. From a practical point of view, the pdf $p_{\bar{e}}$ could be used for choosing the threshold values for the adaptive resampling. The limiting distribution obtained for $N \to \infty$,

$$p_{\bar{e}}(e) = \lim_{N \to \infty} p_{\bar{e}}(e),$$

is also theoretically interesting, since it can characterize the function $E_N$. However, it is not straightforward to obtain $p_{\bar{e}}(e)$ analytically. In Section 6, we approximate different limiting pdf’s $p_{\bar{e}}(e)$ of different G-ESS functions via numerical simulation.

**6. Simulations**

6.1. 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 $\mathbf{w}$ uniformly distributed in the unit simplex $S_N$, and then we compute the corresponding ESS values (e.g., using the procedure described in [8]). We generate 2000 independent random vectors $\mathbf{w}$ uniformly distributed in the unit simplex $S_N \subset \mathbb{R}^N$. After that we evaluate the different proper and stable G-ESS functions (summarized in Table 4) at each drawn vector $\mathbf{w}$. The resulting histograms of the rate ESS/N obtained by the different functions are depicted in Fig. 2. Figs. 2(a)–(c) correspond to $N=50$, whereas (b)–(d) correspond to $N=1000$. Figs. 2(a)–(c) show the histograms of the rate corresponding $D_2^{(\infty)}$, $P_2^{(i)}$, $S_2^{(i)}$, whereas Figs. 2(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 5.

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

\[
\mathbb{E}(\mathbf{w}) \leq \epsilon N,
\]

where \( 0 \leq \epsilon \leq 1 \). Namely, the information provided by Table 5 can be useful for choosing \( \epsilon \), depending on the used G-ESS function. For instance, Doucet et al. [12, Section 3.5] suggest us to use \( \epsilon = \frac{1}{2} \) for \( P_U^2 \). This suggestion can be explained considering the mean of the ESS values of \( P_U^2 \), which is \( \approx 0.5 \). Moreover, the standard deviation can help us to understand the capability of each formula in differentiating different vectors \( \mathbf{w} \). The greatest standard deviation for each \( N \) is highlighted with boldface. In this sense, \( D_{\sigma_0}^2 \) seems the most “discriminative” for large values of \( N \), whereas \( P_U^2 \) seems the more convenient for small values of \( N \) (however, other studies can suggest the opposite; see below).

### 6.2. Approximation of the theoretical ESS definition

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

\[
\text{ESS}_{\text{var}}(h) = N \frac{\text{var}[\tilde{I}(h)]}{\text{var}[I(h)]},
\]

(37)

where we stress the dependence on the choice of the integrand function \( h \). As also discussed in Appendix A, a more convenient definition for small values of \( N \) is

\[
\text{ESS}_{\text{MSE}}(h) = N \frac{\text{MSE}[\tilde{I}(h)]}{\text{MSE}[I(h)]} = N \frac{\text{var}[\tilde{I}(h)]}{\text{MSE}[I(h)]},
\]

(38)

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 \( \tilde{I} \) is virtually zero. In this section, we compute approximately via Monte Carlo the theoretical definitions \( \text{ESS}_{\text{var}}(h) \), \( \text{ESS}_{\text{MSE}}(h) \), and compare with the values obtained with different G-ESS functions. More specifically, we consider a univariate standard Gaussian density as target pdf,

\[
\pi(x) = N(x; 0, 1),
\]

and also a Gaussian proposal pdf,

\[
q(x) = N(x; \mu_p, \sigma_p^2),
\]

(40)

with mean \( \mu_p \) and variance \( \sigma_p^2 \). Furthermore, we consider different 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, \( q(x) \equiv \delta(x) \). As \( \mu_p \) increases, the proposal becomes more different from \( \pi \). We consider the estimation of the expected value of the random variable \( X \sim \delta(x) \), i.e., we set \( h(x) = x \) in the integral of Eq. (1).

**S2** In this case, we set \( \mu_p = 1 \) and consider \( \sigma_p \in \{0.23, 4\} \). We set \( h(x) = x \).

**S3** We fix \( \sigma_p = 1 \) and \( \mu_p \in \{0.3, 0.5, 1, 1.5\} \) and vary the number of samples \( N \). We consider again \( h(x) = x \).

**S4** In order to analyze the dependence on the choice of \( h(x) \) of the theoretical definition (37) and of the numerical results, we consider \( h(x) = x^r, r = 1, ..., R = 10 \). More specifically, we define the averaged ESS (A-ESS) value,

\[
A - \text{ESS} = \frac{1}{R} \sum_{r=1}^{R} \text{ESS}_{\text{var}}(h_r),
\]

(41)

where \( \text{ESS}_{\text{var}}(h_r) \) is given in Eq. (37). First, we set \( \sigma_p = 1 \), \( N \in \{1000, 5000\} \), and vary \( \mu_p \in \{0, 2\} \), as in the setting S1, but we also compute A-ESS in Eq. (41). Then, we set \( \sigma_p = 1 \), \( \mu_p = \{0.3, 1\} \), and vary \( N \), similarly as S3.

In the first two cases, we test \( N \in \{5, 1000\} \). Fig. 3 shows the theoretical ESS curves (approximated via simulations) and the curves corresponding to the proper and stable G-ESS formulas (averaged over \( 10^3 \) independent runs), for the experiment settings S1 and S2. For \( N = 1000 \), the difference between \( \text{ESS}_{\text{var}}(h) \) and \( \text{ESS}_{\text{MSE}}(h) \) is negligible, so that we only show \( \text{ESS}_{\text{var}}(h) \). For \( N = 5 \) and S1 we show both curves of \( \text{ESS}_{\text{var}}(h) \) and \( \text{ESS}_{\text{MSE}}(h) \), whereas for \( N = 5 \) and S2 we only provide \( \text{ESS}_{\text{MSE}}(h) \) since the bias is big for small values of \( \sigma_p \) so that it is difficult to obtain reasonable and meaningful values of \( \text{ESS}_{\text{var}}(h) \). Figs. 4 and 5 provide the results of the experiment setting S3 and S4, respectively. Note that, for simplicity, in Fig. 5 we only show the results of \( D_{\sigma_0}^2 \), \( p_U^2 \) and Ginii, jointly with the theoretical ones, \( \text{ESS}_{\text{var}}(h) \) and A-ESS.

In the setting S1 with \( N = 5 \) shown Fig. 3(a), first of all we
observe that ESSvar(h) and ESSSSE(h) are very close when μp ≈ 0 (i.e., g(x) ≈ 8(h)) but they differ substantially when the bias increases. In this case, the G-ESS function Gini(w) provides the closest values to ESSvar(h), in general. Moreover, P2(w) and D1(w) also provide good approximations of ESSvar(h). Note that ESSvar(h) is always contained between D3(w) and P2(w). In the case S1 with N = 1000 shown Fig. 3(b), the formula P2(w) provides the closest curve to ESSvar(h). The G-ESS function D3(w) gives a good approximation when μp increases, i.e., the scenario becomes worse from a Monte Carlo point of view. The G-ESS function Gini(w) provides the best approximation when μp ∈ [0, 0.5]. Again, ESSvar(h) is always contained between D3(w) and P2(w).

In the second scenario S2 with N = 5 shown Fig. 3(c), all G-ESS functions are not able to reproduce conveniently the shape of ESSMSE(h). Around to the optimal value of σ0, Gini(w) and P2(w) provide the best approximation of ESSMSE(h). For the rest of value of σ0, D1(w) provides the closest results. In the second setting S2 with N = 1000 shown Fig. 3(d), P2(w) seems to emulate better the evolution of ESSvar(h). However, D3(w) provides the closest results for small values of σp.

In the experiment setting S3 (see Fig. 4), we observe that the behavior of the different G-ESS functions as N grows. When μp = 0.3 and μp = 0.5, the function Gini(w) provides the best approximation of the theoretical definition, i.e., ESSvar(h). In particular, with μp = 0.3, Gini(w) seems to approximate precisely the evolution of ESSvar(h). As the proposal differs more to the shape of the target, i.e., for μp = 1 and μp = 1.5, D3(w) becomes the best option. With μp = 1.5, D3(w) reproduces closely the evolution of ESSvar(h). In these last two cases, μp = 1 and μp = 1.5, P2(w) provides also good performance. We conclude that, in this setup, when the proposal and the target substantially differ, D3(w) provides the best results. Roughly speaking, when the shape of proposal is closer to the shape of target, the function Gini(w) provides also good results. Moreover, Gini(w) seems to perform better than P2(w) when the number of particles N is small. In intermediate situations, P2(w) seems to be a good compromise. Finally, in the last setting S4, we can observe (see Fig. 5) that A-ESS in Eq. (41) is in general smaller than ESSvar(h) (which considers only h(x) = x). In these experiments, the G-ESS function D3(w) is the closest approximation of A-ESS.

6.3. Adaptive resampling in particle filtering

In this example, we apply P2(w) and D3(w) within a particle filter in order to decide adaptively when performing a resampling step. Specifically, we consider a stochastic volatility model where the hidden state x follows an AR(1) process and represents the log-volatility [19] of a financial time series at time t ∈ ℤ. The equations of the model are given by

\[ \begin{align*}
  x_t &= a x_{t-1} + u_t, \\
  y_t &= \exp \left( \frac{x_t}{2} \right),
\end{align*} \]

where a = 0.99 is the AR parameter, and ut and vt denote independent zero-mean Gaussian random variables with variances a2 = 1 and a2 = 0.5, respectively. Note that vt is a multiplicative noise. For the sake of simplicity, we implement a standard particle filter (PF) [9,10,17] using as propagation equation of the particles exactly the AR(1) process, i.e., the particles x1:N is propagated as x1:N ∼ DP(x1:N−1, i) = 1, ..., N is the particle index. We set T = 3000 and N = 1000 number of particles. The resampling is performed adaptively, only a certain iterations, T = {τ1, ..., τT},

where r = 0·T (clearly, r varies in each run). More specifically, denoting as w = [w1, ..., wN] at a specific PF iteration t, the conditions for applying the resampling are

\[ P2(w) ≤ εN, \quad D3(w) ≤ εN, \]

respectively, where ε ∈ [0, 1], i = 1, 2 are a constant threshold values (with ε1 = 0, no resampling is performed; with ε2 = 1, the resampling is applied at each iteration).

Let us denote as T1 = {τ1, ..., τT} and T2 = {τ1, ..., τT} the set of resampling instants obtained by P2(w) and D3(w), respectively (t1 = τT1 and t2 = τT2). Since D3(w) ≥ P2(w) for all w ∈ S, and if ε1 = ε2, using D3(w) we apply more resampling steps than when P2(w) is used, i.e., τ2 ≥ τ1 if ε1 = ε2. However, an equal resampling rate R, i.e., the ratio of the averaged number of the performed resampling steps over T,

\[ R = \frac{\text{# Resampling}}{T} = \frac{1}{T} E[R], \]

(44)

can be obtained using different threshold values ε1 and ε2 for P2(w) and D3(w). In our case, for obtaining the same resampling rate we need that ε1 ≥ ε2, as shown in Fig. 6(a). Note that 0 ≤ R ≤ 1.

Goal: Given a resampling rate R, our purpose is to discriminate which G-ESS function, between P2(w) and D3(w), selects the better iteration indices τ*′s for applying the resampling steps, i.e., when it is more adequate to apply resampling in order to improve the performance.

Results: We test 100 different values of ε1 and ε2 (we have considered a thin grid of values from 0 to 1 with width 0.01, for both). For each value of εi, i = 1, 2, we run 500 independent simulations of the PF for inferring the sequence xT obtained, generating a sequence of observations yT generated according to the model in Eq. (42). Hence, we compute the Mean Square Error (MSE) in the estimation of xT obtained by the PF, in each run. Moreover, for each value of εi, i = 1, 2, we calculate the resampling rate R (averaged over the 500 runs). Then, we can plot two curves of averaged MSE versus the resampling rate R, corresponding to P2(w) and D3(w). In this way, we can compare the performance of the PF using the same resampling rate R but obtained with different G-ESS functions, P2(w) and D3(w). The results are shown in Fig. 6(b) in log-log-scale. We see that, for a given resampling rate R, the G-ESS function D3(w) always provides a smaller MSE w.l.o.g. P2(w). This confirms that, at least in certain scenarios, D3(w) is a good measure of ESS and it is a valid alternative for P2(w). Furthermore, the range of useful values of ε in P2(w) is smaller than in D3(w) as shown in Fig. 6(a).

7. Conclusions

In this work, we have proposed new classes of alternative ESS approximations for importance sampling, discussing and testing them from a theoretical and practical point of view. Indeed the novel ESS expressions, jointly with other formulas already presented in the literature, have been classified according to five theoretical requirements presented in this work. This classification has allowed to select six different ESS functions which satisfy all these necessary conditions. Then, we have tested them by numerical simulations. Some of them, such as P2(w) and Gini(w) present interesting features and some benefit, compared to the standard ESS formula P2(w). When the proposal pdf differs substantially to the target density, D3(w) provides the best approximations. When the proposal is close to the target, the function Gini(w) provides also good results. Moreover, Gini(w) seems to perform better than P2(w) when the number of particles N is small. In intermediate scenarios, P2(w) can also be considered a
good compromise. Furthermore, \(D_{m}^{\\nu}(w)\) behaves as a “lower bound” for the theoretical ESS definition, as shown in the numerical simulations. The simulation study also provides some useful value for choosing the threshold in an adaptive resampling context. For instance, the results in Table 5 suggest to use of \(c \geq \frac{1}{2}\) for \(P_{N}^{(2)}(w)\) (as already noted in [12, Section 3.5]) and \(\hat{G}_{m}(w)\), or \(c \geq 0.11\) for \(\hat{D}_{m}^{\\nu}(w)\), in the resampling condition \(E_{m}(w) \leq c N\). We have also tested \(D_{m}^{\\nu}(w)\) and \(P_{N}^{(2)}(w)\) within a particle filter for tracking a stochastic volatility variable. The application of G-ESS function \(D_{m}^{\\nu}(w)\) has provided smaller MSE in estimation w.r.t. \(P_{N}^{(2)}(w)\), considering equal resampling rates (i.e., the number of the performed resampling steps over the total number of iterations of the filter).

### Acknowledgements

The authors would like 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).

### Appendix A. Analysis of the theoretical derivation of \(P_{N}^{(2)}\)

In the following, we summarize the derivation of \(P_{N}^{(2)}\), that can be partially found in [21] and [23, Section 2.5], stressing the multiple approximations and assumptions:

1. In Section 2, the ESS has been conceptually defined as the ratio between the performance of two estimators, the \(\hat{I}\), where samples are drawn from the target, and \(I\), the self-normalized IS estimator. The definition \(\text{ESS} = N \frac{\text{var}(\hat{I})}{\text{MSE}(\hat{I})}\) does not take in account the bias of \(\hat{I}\) (which can be significant for small \(N\)). Therefore, a more complete definition is

\[
\text{ESS} = \frac{\text{MSE}(\hat{I})}{\text{MSE}(I)} = N \frac{\text{var}(I)}{\text{MSE}(I)},
\]

where we have considered the Mean Square Error (MSE) and we have taken into account the bias of \(\hat{I}\). In [21], the derivation starts with the definition in Eq. (5) justifying that “the bias is of the order of \(1/N\) and can be ignored for large \(N\)” and that \(\hat{I}\) is unbiased. Indeed, roughly speaking, the squared bias is typically of order \(N^{-2}\), which is negligible compared to the variance which is of order \(N^{-1}\). Nevertheless, \(P_{N}^{(2)}\) is employed regardless the \(N\). Then, the ratio of variances overestimates the theoretical value \(\text{ESS} = \frac{\text{MSE}(I)}{\text{MSE}(\hat{I})}\).

2. In the derivation of [21], all the samples are considered to be i.i.d. from a single proposal, i.e. \(x_{n} \sim q(x_{n})\), for \(n = 1, ..., N\). Nevertheless, \(P_{N}^{(2)}\) is also used in algorithms which employ multiple proposal pdfs under many different weighting strategies [13].

3. A first delta method is first applied in order to approximate \(\text{var}(\hat{I})\) in [21, Eq. (6)].

4. The second delta method is applied again to approximate the expectation \(E_{n}(w(x, y)| x, y)\) in [21, Eq. (9)], where \(w(x, y) = \frac{a(x, y)}{N}\).

5. In the whole derivation, the target is assumed to be normalized (see [21] and [23, Section 2.5]). This is a strong assumption that very rarely occurs in practical scenarios. If this were not the case, the normalizing constant would appear in [21, Eqs. (7)–(9)], and therefore also in [21, Eqs. (11)–(12)]. As a consequence of the normalized constant assumption, the ESS is approximated as

\[
\text{ESS} \approx \frac{N}{1 + \text{var}(w(x))}.
\]

(A.2)

Since \(w\) is the unnormalized weights, different scaled version of the target would yield different approximation of the ESS. In order to overcome this problem, it has been proposed (see [35] and the further explanation in [34]) to modify the approximation of Eq. (A.2) by

\[
\text{ESS} \approx \frac{N}{1 + \frac{\text{var}(w(x))}{N^2}} = \frac{N}{1 + \text{CV}^2}.
\]

(A.3)

where \(\text{CV}\) represents the coefficient of variation (also as relative standard deviation) defined as the ratio of the standard deviation \(\sqrt{\text{var}(w(x))}\) and the mean \(E_{n}(w(x)) = Z\) [23]. The well-known \(P_{N}^{(2)}\) can be derived as an empirical approximation of Eq. (A.3),

\[
P_{N}^{(2)} \approx \frac{\frac{1}{N} \sum_{n=1}^{N} w_{n}^{2} - \left(\frac{1}{N} \sum_{n=1}^{N} w_{n}\right)^{2}}{\left(\frac{1}{N} \sum_{n=1}^{N} w_{n}\right)^{2}}
\]

(A.4)

\[
P_{N}^{(2)} = \frac{N}{\frac{1}{N} \sum_{n=1}^{N} w_{n}^{2}}.
\]

(A.5)

\[
P_{N}^{(2)} \approx \frac{1}{\sum_{n=1}^{N} w_{n}}.
\]

(A.6)

Nevertheless, if the target distribution is not assumed to be normalized, the approximation of Eq. (A.3) is no longer valid. In other words, the metric \(P_{N}^{(2)}\) is approximated with the assumption of \(Z=1\) in the whole derivation, except in the last step where the \(Z\) is re-incorporated.

Consequences: One consequence of these approximations is

![Fig. A1. Graphical representation of the loss of statistical information normalizing the weights and ignoring the values of the particles (N=3).](image-url)
that, given the values of \( w_i \)'s, the final formula \( P_N^{(2)} \) does not depend directly on the particles \( x_i \), \( n = 1, \ldots, N \), which is obviously a drawback since we are trying to measure the effective sample size of the set of weighted particles (see Fig. A1).\(^4\) Moreover, \( P_N^{(2)} \) is independent from the function \( h \), whereas the theoretical definition \( ESS = N \frac{\text{var}(\hat{I})}{\text{var}(I)} \) involves \( h \). Finally, the inequalities \( 1 \leq P_N^{(2)} \leq N \) always hold, 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 \( ESS = N \frac{\text{var}(\hat{I})}{\text{var}(I)} \). Indeed, by this general definition of \( ESS \), we have \( 0 \leq ESS \leq B, \quad B \geq N \), i.e., namely \( ESS \) can be less than 1, when \( \text{var}(\hat{I}) > \text{var}(I) \), and even greater than \( N \), when \( \text{var}(\hat{I}) < \text{var}(I) \): this case occurs when negative correlation is induced among the generated samples \([13,33]\), Fig. A1 shows the progressive loss of information, first normalizing the weights and then removing the information related to the position of the particles.

**Appendix B. The optimism of \( P_N^{(0)}(\w) \)**

Here, we analyze the behavior of \( P_N^{(0)}(\w) \) in two extreme cases. If all the samples are drawn directly from the target distribution all the weights \( w_i \) are equal, so that \( w_i = \frac{1}{N}, \quad n = 1, \ldots, N \), the vector with equal components \( \w = \frac{1}{N} \cdot 1 \cdot \ldots \cdot N \), is denoted

\[
\w = \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \ldots & \frac{1}{N} \end{bmatrix},
\]

(B.1)

Note that the converse is not always true: namely the scenario \( w_i = \frac{1}{N}, \quad n = 1, \ldots, N \), could occur even if the proposal density is different from the target. Hence, in this case, we can assert \( ESS \leq N \) (considering independent, non-negative correlated, samples). The other extreme case is \( \w^{(j)} = [w_1 = 0, \ldots, w_j = 1, \ldots, w_N = 0] \).

(i.e., \( w_j = 1 \) and \( w_0 = 0 \) (it can occur only if \( \pi(x_0) = 0 \), for \( n \neq j \) with \( j \in \{1, \ldots, N\} \)). The best possible scenario, in this case, is that the \( j \)-th sample (associate to the weight \( w_j = 1 \)) has been generated exactly from \( \pi(x) \) (hence, with effective sample size equal to 1). Thus, in this case, one can consider \( ESS \leq 1 \). The function \( P_N^{(0)}(\w) \) employ an optimistic approach for the two extreme cases previously described above:

\[
P_N^{(0)}(\w^*) = N, \quad (B.3)
\]

\[
P_N^{(0)}(\w^{(j)}) = 1, \quad \forall j \in \{1, \ldots, N\}. \quad (B.4)
\]

Moreover, considering a vector of type

\[
\w = \begin{bmatrix} 0 & 1 & \frac{1}{C} & \ldots & 0 & 0 & \frac{1}{C} & \ldots & 0 \end{bmatrix},
\]

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

\[
P_N^{(0)}(\w) = C. \quad (B.5)
\]

Fig. B1 summarizes graphically these cases. This approach can appear as a limitation given the previous observations but, using

**Appendix C. G-ESS functions induced by non-Euclidean distances**

In Section 3, we have pointed out the relationship between \( P_N^{(2)} \) and the \( L_2 \) distance between \( \w \) and \( \w^* \). Here, we derive other G-ESS functions given based on non-Euclidean distances.

**Distance \( L_1 \):** We derive a G-ESS function denoted as \( Q_N(\w) \), induced by the \( L_1 \) distance. Let us define two disjoint sets of weights

\[
\{\w_1, \ldots, \w_{N^+}\} = \{\text{all } \w_i : \quad \w_i \geq 1/N, \quad \forall \, n = 1, \ldots, N\}, \quad \text{ (C.1)}
\]

\[
\{\w_1, \ldots, \w_{N^-}\} = \{\text{all } \w_i : \quad \w_i < 1/N, \quad \forall \, n = 1, \ldots, N\}, \quad \text{ (C.2)}
\]

where \( N^+ = \#\{\w_i, \ldots, \w_{N^+}\} \) and \( N^- = \#\{\w_1, \ldots, \w_{N^-}\} \). Clearly, \( N^+ + N^- = N \) and \( \sum_{i=1}^{N^+} \w_i^+ + \sum_{i=1}^{N^-} \w_i^- = 1 \). Thus, we can write

\[
\| \w - \w^* \|_1 = \sum_{i=1}^{N} \w_i - \frac{1}{N} = N^+ \left( \w_i - \frac{1}{N} \right) + N^- \left( \frac{1}{N} - \w_i^- \right) = \sum_{i=1}^{N^+} \w_i^+ - \sum_{i=1}^{N^-} \w_i^- = \frac{N^+ - N^-}{N}, \quad \text{ (C.3)}
\]

and replacing the relationships \( \sum_{i=1}^{N^+} \w_i^+ = 1 - \sum_{i=1}^{N^-} \w_i^- \) and \( N^+ = N - N^- \)

\[
\| \w - \w^* \|_1 = 2 \left( \frac{N^+ - N^-}{N} \right) = 2 \left( N \sum_{i=1}^{N^+} \w_i^+ - N^- \right) = 2 \left[ \frac{N - Q_N(\w)}{N} \right], \quad \text{ (C.4)}
\]

where

\[
Q_N(\w) = - N \sum_{i=1}^{N^+} \w_i^+ + N^+ + N. \quad \text{ (C.5)}
\]

Note that \( 1 \leq Q_N(\w) \leq N \), with \( Q_N(\w^*) = N \) and \( Q_N(\w^{(j)}) = 1 \) for all \( j \in \{1, \ldots, N\} \). Maximizing \( Q_N \) is equivalent to minimizing the \( L_1 \) distance between the pmf \( \w \) and the discrete uniform pmf \( \w^* \). We remark that this is only one of the possible ESS functions induced by the \( L_1 \) distance. We choose \( Q_N \) since it is proper and stable.

**Norm \( L_0 \):** Interesting G-ESS expressions can be also obtained
considering also the distance of the vector \( \mathbf{w} \) with respect to the null vector containing all zeros as entries (i.e., the norm of \( \mathbf{w} \)). For instance, based on the Hamming distance \([7]\), i.e., we have
\[
V_N^{(s)}(\mathbf{w}) = N - N_0, \quad (C.6)
\]
where \( N_0 \) is the number of zeros in \( \mathbf{w} \), i.e.,
\[
N_0 = \#(w_n = 0, \quad \forall \ n = 1, \ldots, N).
\]

Observe that \( 1 \leq V_N^{(s)}(\mathbf{w}) \leq N \) and \( V_N^{(0)}(\mathbf{w}) = N \) and \( V_N^{(s)}(\mathbf{w}^i) = 1 \) for all \( i \in \{1, \ldots, N\} \).

**Norm \( L_\infty \):** Other kind of norms can suggest other suitable ESS formulas. For instance,
\[
\| \mathbf{w} \|_\infty = \max \{ |w_0|, \ldots, |w_N| \} = \frac{1}{D_N(\mathbf{w})},
\]
where
\[
D_N^{(\infty)}(\mathbf{w}) = \max \{ |w_0|, \ldots, |w_N| \}.
\] (C.8)

This G-ESS function has also been recently considered in [18].

**Appendix D. Derivation of generalized ess families**

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 \( f(\mathbf{w}): \mathbb{R}^N \to \mathbb{R} \), which satisfies the following properties:
1. \( f(\mathbf{w}) \) is a quasi-concave or a quasi-convex function, with a minimum or a maximum (respectively) at \( \mathbf{w} = \left\{ \frac{1}{N}, \ldots, \frac{1}{N} \right\} \).
2. \( f(\mathbf{w}) \) is symmetric in the sense of Eq. \((11)\).
3. Considering the vertices of the unit simplex \( \mathbf{w}^i = \delta(i) \) in Eq. \((B.2)\), then we also assume

\[
f(\mathbf{w}^i) = c,
\]

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

Let also consider the function \( a f(\mathbf{w}) + b \) obtained as a linear transformation of \( f(\mathbf{w}) \) when \( a, b \in \mathbb{R} \) are two constants. Note that, we can always set \( a > 0 \) if \( f(\mathbf{w}) \) is quasi-concave, or \( a < 0 \) if \( f(\mathbf{w}) \) is quasi-convex, in order to obtain \( g(\mathbf{w}) \) is always quasi-concave. Hence, we can define the G-ESS function as
\[
E_{af}(\mathbf{w}) = \frac{1}{af(\mathbf{w}) + b}, \quad \text{or} \quad E_{ag}(\mathbf{w}) = af(\mathbf{w}) + b.
\]

In order to fulfill the properties 2 and 3 in Section 4, recalling \( \mathbf{w} = \left\{ \frac{1}{N}, \ldots, \frac{1}{N} \right\} \) and \( \mathbf{w}^i = \delta(i) \), 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 \( b \),
\[
\begin{cases}
a f(\mathbf{w}) + b = 1 \\
a f(\mathbf{w}^i) + b = 1, \quad \forall \ i \in \{1, \ldots, N\}.
\end{cases}
\] (D.1)

or
\[
\begin{cases}
a f(\mathbf{w}) + b = N \\
a f(\mathbf{w}^i) + b = 1, \quad \forall \ i \in \{1, \ldots, N\}.
\end{cases}
\]

respectively. Note that they are both linear with respect to \( a \) and \( b \). Moreover, since \( f(\mathbf{w}^i) = c \) for all \( i \in \{1, \ldots, N\} \), the system above is reduced to a \( 2 \times 2 \) linear system with solution
\[
\begin{cases}
a = \frac{N - 1}{N[f(\mathbf{w}^i) - f(\mathbf{w})]} \\
b = \frac{f(\mathbf{w}^i) - f(\mathbf{w})}{N[f(\mathbf{w}^i) - f(\mathbf{w})]}.
\end{cases}
\] (D.2)

and
\[
\begin{cases}
a = \frac{N - 1}{f(\mathbf{w}) - f(\mathbf{w}^i)} \\
b = \frac{f(\mathbf{w}) - f(\mathbf{w}^i)}{f(\mathbf{w}) - f(\mathbf{w}^i)}.
\end{cases}
\] (D.3)

respectively. Below, we derive some special cases of the families \( P_N^{(l)} \), \( D_N^{(l)} \), \( V_N^{(l)} \), and \( S_N^{(l)} \) defined in Section 5 and obtained used the procedure above. In these families, we have \( f(\mathbf{w}) = \sum_{n=1}^{N} w_n^r \) for \( P_N^{(l)} \) and \( V_N^{(l)} \), and \( f(\mathbf{w}) = \left[ \sum_{n=1}^{N} w_n^r \right]^\frac{1}{r} \) for \( D_N^{(l)} \) and \( S_N^{(l)} \).

**D.1. Special cases of \( P_N^{(l)}(\mathbf{w}) \)**

In the following, we analyze some special cases of the family \( P_N^{(l)}(\mathbf{w}) \) in Eq. \((17)\):

**Case \( r = 0 \):** In this case, the constants in Table 2 reach the values \( a_r = d_0 = -\frac{1}{N} \) and \( b_r = b_0 = \frac{N + 1}{N} \). 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 \( \mathbf{w} \) then \( f_0(\mathbf{w}) = N \) and \( P_N^{(0)}(\mathbf{w}) = \frac{1}{N_0 + \delta} \) is \( N \), whereas if \( \mathbf{w} \) contains \( N_2 \) zeros, we have \( f_0(\mathbf{w}) = N - N_2 \) and
\[
P_N^{(0)}(\mathbf{w}) = \frac{N}{N_2 + 1}.
\] (D.5)

where we recall that \( N_2 \) is the number of zero within \( \mathbf{w} \). Note that, clearly, \( P_N^{(0)}(\mathbf{w}) = 1 \) for all \( i \in \{1, \ldots, N\} \), since \( N_2 = N - 1 \).

**Case \( r = 1 \):** In this case, \( a_r = \pm \infty, b_r = \mp \infty \), when \( r = 1 \). Since \( f_r(\mathbf{w}) \to 1 \) if \( r \to 1 \), we have an indeterminant form for \( e_r(\mathbf{w}) = a_r + b_r \), of type \( \infty - \infty \). Note that the limit
\[
\lim_{r \to 1} P_N^{(l)}(\mathbf{w}) = \lim_{r \to 1} \frac{N^{l(2-r)} - N}{(1 - N) \sum_{n=1}^{N} w_n^r + N^{(2-r)} - 1}
\]
presents an indeterminant form of type \( \frac{\infty}{\infty} \). Hence, using the L'Hôpital's rule [20], i.e., deriving both numerator and denominator w.r.t. \( r \) and computing the limit, we obtain
\[
P_N^{(1)}(\mathbf{w}) = \lim_{r \to 1} \frac{-N^{l(2-r)} \log(N)}{(1 - N) \sum_{n=1}^{N} w_n^r \log(w_n) - N \log(N) + (N - 1) \sum_{n=1}^{N} \log(w_n) - N \log(N) + (N - 1) \log(e) - N \log(N) + (N - 1) H(\mathbf{w}) - N \log(N) + (N - 1) H(\mathbf{w})}
\] (D.7)

where we have denoted as \( H(\mathbf{w}) = -\sum_{n=1}^{N} \log(w_n) \) the discrete entropy of the pmf \( w_n, n = 1, \ldots, N \). Observe that \( H(\mathbf{w}) = \log_N \) then \( P_N^{(1)}(\mathbf{w}) = -N \log(N) \) is \( N \), whereas \( H(\mathbf{w}^i) = 0 \) (considering \( 0 \log(0) = 0 \)).

**Case \( r = 2 \):** In this case, \( a_2 = 1 \) and \( b_2 = 0 \), hence we obtain
\[ P_N^{(2)}(\mathbf{w}) = \frac{1}{\sum_{i=1}^{N} (\mathbf{w}_i)^2}. \]

Case \( r \to \infty \): We have \( a_i \to a = \frac{N-1}{N} \) and \( b_i \to b = \frac{1}{N} \). If \( \mathbf{w} \neq \mathbf{w}^{(0)} \) for all possible \( i \in \{1, ..., N\} \), then we have \( \lim_{r \to \infty} f_i(\mathbf{w}) = 0 \) (since \( 0 < \bar{w}_i < 1 \). in this case) and \( P_N^{(1)}(\mathbf{w}) = \frac{1}{N}. \) Otherwise, if \( \mathbf{w} = \mathbf{w}^{(0)} \) for some \( i \in \{1, ..., N\} \), then \( \lim_{r \to \infty} f_i(\mathbf{w}) = 1 \) (where we have considered \( \lim_{r \to \infty} \mathbf{0} = 0 \) and \( \lim_{r \to \infty} \mathbf{T} = 1 \)) and \( P_N^{(1)}(\mathbf{w}) = \frac{1}{a_i + b_i} = 1 \). We can summarize both scenarios as
\[ P_N^{(1)}(\mathbf{w}) = \begin{cases} N, & \text{if } \mathbf{w} \neq \mathbf{w}^{(0)}, \forall i \in \{1, ..., N\}, \\ 1, & \text{if } \mathbf{w} = \mathbf{w}^{(0)}, \forall i \in \{1, ..., N\}. \end{cases} \] (D.8)

### D.2. Special cases of \( D_N^{(1)}(\mathbf{w}) \)

Below, we analyze some special cases of the family \( D_N^{(1)}(\mathbf{w}) \) in Eq. (18):

- **Case \( r \to 0 \):** The coefficients of this family given in Table 2 reach the values \( a_i = a = \frac{1}{N} \) and \( b_i = b = 1 \). In this case, \( f_i(\mathbf{w}) = 1 \) (considering again \( \mathbf{0} = 0 \) and \( \mathbf{T} = 1 \)). Whereas, when \( \mathbf{w} \) is not a vector, i.e., \( \mathbf{w} \neq \mathbf{w}^{(0)} \), then
\[ \lim_{r \to 0} f_i(\mathbf{w}) = \lim_{r \to 0} \left[ \sum_{i=1}^{N} (\bar{w}_i)^r \right]^{1/r} = \infty. \]

so that \( \lim_{r \to 0} \alpha_i f_i(\mathbf{w}) \) has the indeterminate form of type \( 0 \times \infty \) that can be converted to \( \frac{0}{0} \) as shown below. We can write
\[ \lim_{r \to 0} \left[ \sum_{i=1}^{N} (\bar{w}_i)^r \right]^{1/r} = (1 - N) \lim_{r \to 0} \left[ \sum_{i=1}^{N} (\bar{w}_i)^r \right]^{1/r}. \] (D.9)

Moreover, when \( r \to 0 \) we have
\[ \left[ \sum_{i=1}^{N} (\bar{w}_i)^r \right]^{1/r} \approx \left[ \frac{\sum_{i=1}^{N} (\bar{w}_i)^r}{N^{1/r}} \right]^{1/r} = \left[ \frac{N}{\sum_{i=1}^{N} (\bar{w}_i)^r} \right]^{1/r}, \]

when \( r \to 0 \).

For \( r \to 0 \), we can also write
\[ (\bar{w}_i)^r = \exp(r \log(\bar{w}_i)) \approx 1 + r \log(\bar{w}_i), \]

where we have used the Taylor expansion of first order of \( \exp(r \log(\bar{w}_i)) \). Replacing \( \bar{w}_i \approx 1 + r \log(\bar{w}_i) \), the sum inside \( \frac{N}{\sum_{i=1}^{N} (\bar{w}_i)^r} \) we obtain
\[ \frac{1}{N} \sum_{i=1}^{N} (\bar{w}_i)^r \approx \frac{1}{N} \sum_{i=1}^{N} \log(\bar{w}_i) = 1 + r \log \prod_{i=1}^{N} \bar{w}_i. \] (D.10)

Moreover, given \( x \in \mathbb{R} \), for \( r \to 0 \) we also have the relationship
\[ (1 + rx)^r \to \exp(x), \]

by definition of exponential function. Replacing the above, for \( r \to 0 \),
\[ \left[ 1 + \frac{\log \left( \prod_{i=1}^{N} \bar{w}_i \right)^{1/r} \right]^{1/r} \to \left( \log \prod_{i=1}^{N} \bar{w}_i \right)^{1/r}. \] (D.15)

Thus, finally we obtain
\[ \lim_{r \to 0} \left[ \sum_{i=1}^{N} (\bar{w}_i)^r \right]^{1/r} = (1 - N^{1/r}) \prod_{i=1}^{N} \bar{w}_i^{1/r}, \]

and
\[ \lim_{r \to 0} D_N^{(1)}(\mathbf{w}) = \frac{1}{(1 - N^{1/r}) \prod_{i=1}^{N} \bar{w}_i^{1/r} + 1}. \] (D.16)

\[
D_N^{(0)}(\mathbf{w}) = \frac{1}{(1 - N) \text{GeoM}(\mathbf{w}) + 1}
\] (D.17)

### D.3. Special cases of \( V_N^{(1)}(\mathbf{w}) \)

In the following, we study some special cases of the family \( V_N^{(1)}(\mathbf{w}) \) in Eq. (19):

- **Case \( r \to 0 \):** The coefficients of this family given in Table 2 are \( a_i = 1 \) and \( b_i = 0 \). If \( \mathbf{w} \) does not contain zeros then \( f_i(\mathbf{w}) = \prod_{i=1}^{N} \bar{w}_i = N \). Otherwise, assuming \( \mathbf{0} = 0 \), if \( \mathbf{w} \) contains \( N \) zeros, we have \( f_i(\mathbf{w}) = \prod_{i=1}^{N} \bar{w}_i = N - N \). Thus, in general, we have
\[ V_N^{(0)}(\mathbf{w}) = N - N. \]

- **Case \( r \to 1 \):** In this, the coefficients \( a_i \) and \( b_i \) diverge. We can consider the limit
\[ \lim_{r \to 1} \left[ \frac{N^{r-1}(N-1)}{1 - N^{r-1}} \prod_{i=1}^{N} \bar{w}_i \right] = (N - 1) \lim_{r \to 1} \left[ \frac{\sum_{i=1}^{N} \bar{w}_i^{r-1} - 1}{1 - N^{r-1}} \right]. \]

where we have a indetermination of type \( \frac{0}{0} \). Using the L'Hôpital's rule [20], i.e., deriving both numerator and denominator w.r.t. \( r \) and computing the limit, we obtain (since \( \log y = \frac{\log x}{\log e} \))
\[ \lim_{r \to 1} \left[ \frac{\sum_{i=1}^{N} \bar{w}_i^{r-1} - 1}{1 - N^{r-1}} \right] = \lim_{r \to 1} \left[ \frac{\sum_{i=1}^{N} \bar{w}_i^{r-1} \log(\bar{w}_i)}{1 - N^{r-1}} \right]. \]

Thus, we can write
\[ \lim_{r \to 1} \left[ 1 + \frac{\log \left( \prod_{i=1}^{N} \bar{w}_i \right)^{r-1}}{\log N} \right]^{1/(r-1)} \to \frac{H(\mathbf{w})}{\log N}. \]

hence, finally,
\[ V_N^{(1)}(\mathbf{w}) = (N - 1) \frac{H(\mathbf{w})}{\log N} + 1. \] (D.19)

- **Case \( r \to \infty \):** The coefficients converge to the values \( a_i = a = 1 - N \) and \( b_i = b = N \). If \( \mathbf{w} \neq \mathbf{w}^{(0)} \) then
\[ f_r(\mathbf{w}) = \sum_{i=1}^{N} (\mathbf{w}_i)^r = 0, \text{ so that } V_N^r(\mathbf{w}) = b_{\infty} = N. \] Otherwise, if \( \mathbf{w} = \mathbf{w}^{(0)} \), since \( 0^0 = 0 \) and considering \( F^r = 1 \), then \( f_r(\mathbf{w}) = \sum_{i=1}^{N} (\mathbf{w}_i)^r = 1, \) so that \( V_N^r(\mathbf{w}) = a_\infty + b_{\infty} = 1. \)

D.4. Special cases of \( S_n^r(\mathbf{w}) \)

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

Case \( r \to 0 \): The coefficients given in Table 2 in this case are \( a_r = b = 0 \) and \( b_r = 1 \). If \( \mathbf{w} \neq \mathbf{w}^{(0)} \), then \( f_r(\mathbf{w}) \to \infty \). Otherwise, if \( \mathbf{w} = \mathbf{w}^{(0)} \), since \( 0^0 = 0 \) and considering \( F^r = 1 \), then \( f_r(\mathbf{w}) \to 1. \) With a procedure similar to \( D_0^r(\mathbf{w}) \), it is possible to show that

\[ S_n^0(\mathbf{w}) = (N^2 - N)\text{GeoM}(\mathbf{w}) + 1. \] (D.20)

Case \( r \to \frac{1}{2} \): We have \( a_{1/2} = 1 \) and \( b_{1/2} = 0. \) Then, in this case,

\[ S_n^{1/2}(\mathbf{w}) = f_{1/2}(\mathbf{w}) = \left( \sum_{i=1}^{N} \sqrt{\mathbf{w}_i} \right)^2. \]

Case \( r \to 1 \): With a similar procedure used for \( V_N^1(\mathbf{w}) \), it is possible to obtain

\[ S_n^1(\mathbf{w}) = (N - 1) \log^2 N + 1. \] (D.21)

Case \( r \to \infty \): In this case, \( a_r = a_\infty = -N, \ b_r = b_\infty = N + 1. \) Moreover, \( f_r(\mathbf{w}) \to \max(\mathbf{w}_1, \ldots, \mathbf{w}_N) \) \cite{28}, so that

\[ S_n^\infty(\mathbf{w}) = (N + 1) - N\max(\mathbf{w}_1, \ldots, \mathbf{w}_N). \] (D.22)

Table 3 summarizes all the special cases.

References