ALTERNATIVE EFFECTIVE SAMPLE SIZE MEASURES FOR IMPORTANCE SAMPLING

L. Martino†, V. Elvira‡, F. Louzada§

† Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, Leganés (Spain).
§ Institute of Mathematical Sciences and Computing, Universidade de São Paulo, São Carlos (Brazil).

ABSTRACT

The Effective Sample Size (ESS) is an important measure of efficiency in the Importance Sampling (IS) technique. A well-known approximation of the theoretical ESS definition, involving the inverse of the sum of the squares of the normalized importance weights, is widely applied in literature. This expression has become an essential piece within Sequential Monte Carlo (SMC) methods, using adaptive resampling procedures. In this work, first we show that this ESS approximation is related to the Euclidean distance between the probability mass function (pmf) described by the normalized weights and the uniform pmf. Then, we derive other possible ESS functions based on different discrepancy measures. In our study, we also include another ESS measure called perplexity, already proposed in literature, that is based on the discrete entropy of the normalized weights. We compare all of them by means of numerical simulations.

Index Terms— Effective Sample Size; Importance Sampling; Perplexity measure; Resampling; Sequential Monte Carlo.

1. INTRODUCTION

The Effective Sample Size (ESS) is a widely used concept for measuring the efficiency of different Monte Carlo methods, such as Markov Chain Monte Carlo (MCMC) [11, 15, 20] and Importance Sampling (IS) techniques [1, 4, 18, 16, 21]. 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 [11, 13] 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 \( \tilde{ESS} \approx \frac{1}{\sum_{n=1}^{N} \bar{w}^2_n} \), which invol-
volves (only) the normalized importance weights \( \bar{w}_n, n = 1, \ldots, N \) [6, 7, 14, 21]. This expression presents different weaknesses since it has been obtained after several approximations of the theoretical definition (see [17] for further details). Another measure called perplexity, involving the discrete entropy [5] of the normalized weights has been also proposed in [2]; see also [21, Chapter 4], [9, Section 3.5].

However, the approximation \( \tilde{ESS} \) is widely used in different Sequential Monte Carlo (SMC) methods (a.k.a., particle filtering algorithms) [7, 8, 6, 12, 19]. A key point for the success of a SMC method is the use of resampling procedures, that are applied for avoiding the particle degeneracy [6, 7]. 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 an approximation of the ESS [6, 16, 21]. We show that \( \tilde{ESS} \) is related to Euclidean distance 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, \( \tilde{ESS} \) provides high values otherwise, when the pmf defined by \( \bar{w}_n \) is concentrated mainly in one weight, \( \tilde{ESS} \) provides small values. We deduce other possible ESS functions based on different distances. We compare them by means of numerical simulations. This analysis shows that (at least) one novel 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 \( \frac{1}{\sum_{n=1}^{N} \bar{w}^2_n} \).

2. EFFECTIVE SAMPLE SIZE FOR IMPORTANCE SAMPLING

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

\[
I = \int_{\mathcal{X}} h(x) \tilde{\pi}(x) dx,
\]

(1)
which 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 \( I \) is

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

(2)

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 \( q(x) \),\(^1\) and then assign a weight to each sample, \( w_n = \frac{\pi(x_n)}{q(x_n)} \), with \( n = 1, \ldots, N \), according to the importance sampling (IS) approach. Defining the normalized weights,

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

then the self-normalized IS estimator is

\[
\bar{I} = \sum_{n=1}^{N} \bar{w}_n h(x_n) \approx I,
\]

(4)

with \( x_n \sim q(x) \), \( n = 1, \ldots, N \). In general, the estimator \( \bar{I} \) is less efficient than \( \tilde{I} \), since the samples are not directly drawn from \( \pi(x) \). In several applications \([6, 7, 12, 19]\), it is necessary to measure the loss of the efficiency using the first equality, and the unnormalized ones in the second equality. An interesting property of the expression (7) is

\[
1 \leq P_N(\bar{w}) \leq N.
\]

(8)

Due to the several approximations which have been applied to obtain the final formula, \( P_N \) does not depend on the particles \( x_n, n = 1, \ldots, N \), which is obviously a drawback (for further considerations see \([17]\)). Another similar measure, called perplexity, has been proposed in literature \([2, 21]\) based only on the normalized importance weights,

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

(9)

where

\[
H(\bar{w}) = -\sum_{n=2}^{N} \bar{w}_n \log \bar{w}_n
\]

is the discrete entropy of the vector \( \bar{w} \) \([5]\).\(^2\) Note that, \( 1 \leq \text{Per}_N(\bar{w}) \leq N \).

3. ESS BASED ON DISCREPANCY MEASURES

Many population Monte Carlo (PMC) \([3, 10]\) or sequential Monte Carlo (SMC) methods \([7, 12]\), employ resampling steps for updating the parameters of the used proposal functions. On the one hand, PMC and SMC suffer the so-called particle degeneracy, i.e., after some iterations only one sample is statistically relevant in terms of importance weights. This problem could be solved by applying resampling procedures. However, on the other hand, the application of resampling yields loss of diversity in the set of samples (and 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 the 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 were drawn directly from the target distribution all the weights \( w_n \) would be equal, so that \( \bar{w}_n = \frac{1}{N} \), \( n = 1, \ldots, N \). We denote

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

(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 \( ESS \leq N \)\(^3\).

\(^1\)We assume that \( q(x) > 0 \) for all \( x \) where \( \pi(x) \neq 0 \), and \( q(x) \) has heavier tails than \( \pi(x) \).

\(^2\)Different ESS approximations involving the discrete entropy can be designed. However, the perplexity satisfies certain important properties \([17]\).

\(^3\)
(considering independent samples). The other extreme case is
\[
\bar{w}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_N = 0],
\]
\(11\)
i.e., \(\bar{w}_j = 1\) and \(\bar{w}_n = 0\) (it occurs only if \(\pi(x_n) = 0\), for \(n \neq j\) with \(j \in \{1, \ldots, N\}\). In general, in this case \(ESS \leq 1\). Both approximations \(P_N\) and \(\text{Per}_N\), based only on the information given by the vector \(\bar{w}\), apply an optimistic approach setting \(ESS = N\) and \(ESS = 1\) in the extreme scenarios described above. The underlying idea behind both formulas \(P_N(\bar{w})\), \(\text{Per}_N(\bar{w})\), is that if the pmf \(\bar{w}_n\), \(n = 1, \ldots, N\), is reasonably close to the discrete uniform pmf \(U\{1, 2, \ldots, N\}\) then the resampling is not needed. Otherwise, the resampling is applied. Below, we show that the formula \(P_N\) is related to the Euclidean distance between these two pdfs. We derive some alternative ESS functions employing other kind of distances between the pmf represented by weights \(\bar{w}_n\) and the discrete uniform pmf. For further information see [17].

**Euclidean distance** \(L_2\). Let us consider the Euclidean distance \(L_2\) between the discrete uniform pmf \(U\{1, 2, \ldots, N\}\) and the pmf given by the normalized weights \(\bar{w}_n\), i.e.,
\[
||\bar{w} - \bar{w}^*||_2 = \sqrt{\sum_{n=1}^{N} \left( \bar{w}_n - \frac{1}{N} \right)^2}
\]
\(12\)
\[
= \sqrt{\sum_{n=1}^{N} \frac{\bar{w}_n^2}{N}} + N \left( \frac{1}{N^2} \right) - 2 \frac{N}{N} \sum_{n=1}^{N} \bar{w}_n
\]
\[
= \sqrt{\sum_{n=1}^{N} \frac{\bar{w}_n^2}{N}} - \frac{1}{N}
\]
\[
= \sqrt{\frac{1}{P_N(\bar{w})} - \frac{1}{N}}.
\]

Therefore, maximizing \(P_N\) is equivalent to minimizing the Euclidean distance between the pmf \(\bar{w}_n\) and the discrete uniform pmf.

**Distance** \(L_1\). Given the previous observations, we can attempt to obtain other suitable ESS formulas, employing other distances. Let us define two disjoint sets of weights
\[
\{\bar{w}_1^+, \ldots, \bar{w}_N^+\} = \{\text{all } \bar{w}_n: \bar{w}_n \geq 1/N, \forall n = 1, \ldots, N\},
\]
\[
\{\bar{w}_1^-, \ldots, \bar{w}_N^-\} = \{\text{all } \bar{w}_n: \bar{w}_n < 1/N, \forall n = 1, \ldots, N\},
\]
where \(N^+ = \#\{\bar{w}_1^+, \ldots, \bar{w}_N^+\}\) and \(N^- = \#\{\bar{w}_1^-, \ldots, \bar{w}_N^-\}\).

Clearly, \(N^- + N^+ = N\) and \(\sum_{i=1}^{N^+} \bar{w}_i^+ + \sum_{i=1}^{N^-} \bar{w}_i^- = 1\).

Considering the \(L_1\) distance, we can write
\[
||\bar{w} - \bar{w}^*||_1 = \sum_{n=1}^{N} \left| \bar{w}_n - \frac{1}{N} \right|
\]
\(13\)
\[
= \sum_{i=1}^{N^+} \left( \bar{w}_i^+ - \frac{1}{N} \right) + \sum_{j=1}^{N^-} \left( \frac{1}{N} - \bar{w}_j \right)
\]
\[
= \sum_{i=1}^{N^+} \bar{w}_i^+ - \sum_{i=1}^{N^-} \bar{w}_i^- + \frac{N^+ - N^-}{N}
\]
\(14\)
and replacing the relationships \(\sum_{i=1}^{N^+} \bar{w}_i^- = 1 - \sum_{i=1}^{N^+} \bar{w}_i^+\) and \(N^- = N - N^+\),
\[
||\bar{w} - \bar{w}^*||_1 = 2 \left[ \sum_{i=1}^{N^+} \bar{w}_i^+ - \frac{N^+}{N} \right],
\]
\(15\)
\[
= 2 \frac{N \sum_{i=1}^{N^+} \bar{w}_i^+ - N^+}{N} = 2 \left[ \frac{N - Q_N(\bar{w})}{N} \right] + 2,
\]
where
\[
\widehat{ESS} = Q_N(\bar{w}) = -N \sum_{i=1}^{N^+} \bar{w}_i^+ + N^+ + N,
\]
\(16\)
Note that \(1 \leq Q_N(\bar{w}) \leq N\), with \(Q_N(\bar{w}^*) = N\) and \(Q_N(\bar{w}^{(i)}) = 1\) for all \(i \in \{1, \ldots, N\}\). Maximizing \(Q_N\) is equivalent to minimizing the \(L_1\) distance between the pmf \(\bar{w}_n\) and the uniform pmf.

**Norm** \(L_\infty\). Considering the distance between the vector \(\bar{w}\) and the vector of null entries (i.e. the norm of \(\bar{w}\)), we can obtain other suitable ESS formulas. For instance, we can consider the norm \(L_\infty\), i.e.,
\[
||\bar{w}||_\infty = \max ||\bar{w}_1, \ldots, \bar{w}_N|| = \frac{1}{D_N(\bar{w})},
\]
\(17\)
where
\[
\widehat{ESS} = D_N(\bar{w}) = \left[ \frac{1}{\max \{\bar{w}_1, \ldots, \bar{w}_N\}} \right],
\]
\(18\)
is another valid ESS measure. We have also \(1 \leq D_N(\bar{w}) \leq N\), with \(D_N(\bar{w}^*) = N\) and \(D_N(\bar{w}^{(i)}) = 1, \forall i \in \{1, \ldots, N\}\).

**4. NUMERICAL SIMULATIONS**

Let us recall the theoretical definition of ESS in Eq. (5),
\[
ESS_{\text{var}} = N \frac{\text{var}_\pi[\bar{I}]}{\text{var}_q[\bar{I}]}.
\]
\(19\)
Since \(\bar{I}\) is biased and its bias is not negligible for small \(N\), a more convenient definition is
\[
ESS_{\text{MSE}} = N \frac{\text{MSE}_\pi[\bar{I}]}{\text{MSE}_q[\bar{I}]} = N \frac{\text{var}_\pi[\bar{I}]}{\text{MSE}_q[\bar{I}]}.
\]
We can conclude that in general when the proposal differs substantially from the target, $D_N$ provides the best results (i.e., closer to the theoretical values), whereas in better scenarios and large $N$, $P_N$ seems to be the best approximations. The ESS function $D_N$ seems to perform better than $P_N$ when the number of particles $N$ is small. The ESS function $Q_N$ provides the best approximation when the proposal is very similar to the target in both cases $N = 5$ and $N = 1000$. In the analyzed scenarios, the perplexity does not seem a suitable approximation of the theoretical ESS values.

5. CONCLUSIONS

In this work, we have introduced alternative ESS functions for sampling algorithms based on IS. They are derived based on discrepancy measures between the multinomial pmf defined by the normalized weights $\bar{w}_n, n = 1, \ldots, N$, and the discrete uniform pmf (or the vector of null entries). We have shown that the standard ESS approximation $P_N$ is related to the Euclidean distance $L_2$. Another measure called perplexity [2, 21] can be included in this class of ESS approximations based on discrepancy measure, since it is based on the discrete entropy of the normalized weights $\bar{w}_n, n = 1, \ldots, N$. We have tested and compared all them by numerical simulations.

At least one of them, $D_N(\bar{w}) = \frac{1}{\max[\bar{w}_1, \ldots, \bar{w}_N]}$, presents interesting features and some benefit, compared to the rest of ESS formulas, when the proposal function differs substantially from the target distribution. Moreover, $D_N(\bar{w})$ seems to behave as a “lower bound” for the theoretical ESS definition, in our simulations.
References


