EFFECTIVE SAMPLE SIZE FOR IMPORTANCE SAMPLING (FUNNY STORY)

L. Martino

jointly with V. Elvira

The Magic of the L'Hospital's Rule...

Approximating integrals

- In Bayesian inference, we often need to compute efficiently integrals involving the (posterior) target pdf $\bar{\pi}(\mathbf{x}) = \frac{1}{Z}\pi(\mathbf{x})$,

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

- We approximate $I(h)$ by Monte Carlo methods.

Monte Carlo - Importance Sampling (IS)

- STANDARD-"IDEAL" CASE: Draw $\mathbf{x}_n \sim \bar{\pi}(\mathbf{x})$, $n = 1, \ldots, N$, and

$$\widehat{I}(h) = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{x}_n) \underset{N \to \infty}{\longrightarrow} I.$$

- In general, it is not possible to draw from $\bar{\pi}(\mathbf{x})$.

Monte Carlo - Importance Sampling (IS)

- STANDARD-"IDEAL" CASE: Draw $\mathbf{x}_n \sim \bar{\pi}(\mathbf{x})$, $n = 1, \dots, N$, and

$$\widehat{I}(h) = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{x}_n) \underset{N\to\infty}{\longrightarrow} I.$$

- In general, it is not possible to draw from $\bar{\pi}(\mathbf{x})$.
- IMPORTANCE SAMPLING (IS): Draw $\mathbf{x}_n \sim q(\mathbf{x})$, $n = 1, \dots, N$,

$$\widetilde{I}(h) = \sum_{n=1}^{N} \bar{w}_n h(\mathbf{x}_n) \underset{N\to\infty}{\longrightarrow} I.$$

where

$$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \qquad \bar{w}_n = \frac{w_n}{\sum_{n=1}^{N} w_n}.$$

Effective Sample Size (ESS)

- Generally, using IS, we lose some efficiency w.r.t. the standard MC case.
- From a *theoretical* and *practical* point of view, it is important to *measure* this loss of efficiency.
- Statistically speaking: do my N weighted samples correspond to E samples (with $E < N$) independently drawn from $\bar{\pi}$?
- A possible math-definition of the Effective Sample Size (ESS) is:

$$E = ESS = N \frac{\text{var}_{\pi}[\widehat{I}]}{\text{var}_{q}[\widetilde{I}]}. \tag{2}$$

See **[Kong92]**.

Effective Sample Size (ESS) - definition drawbacks/observations

- The definition depends on $h(\mathbf{x})$:

$$ESS(h) = N \frac{\text{var}_\pi[\widehat{I}(h)]}{\text{var}_q[\widetilde{I}(h)]}.$$

- A more complete definition should be:

$$ESS = N \frac{\text{MSE}_\pi[\widehat{I}]}{\text{MSE}_q[\widetilde{I}]} = N \frac{\text{var}_\pi[\widehat{I}]}{\text{MSE}_q[\widetilde{I}]}. \quad (3)$$

Effective Sample Size in practice

- However, the theoretical formula is "useless" from a practical point of view.

$$ESS = N\frac{\mathrm{var}_{\pi}[\widehat{I}]}{\mathrm{var}_{q}[\widetilde{I}]} \xrightarrow{Kong92} \widehat{ESS} = ?$$

- Try to find something that we can easily compute.

Effective Sample Size in practice

- After several approximations and assumptions, one can obtain

$$ESS \approx \widehat{ESS} = P_N^{(2)}(\bar{\mathbf{w}}) = \frac{1}{\sum_{n=1}^{N}(\bar{w}_n)^2}, \qquad (4)$$

where $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$ is the vector of normalized weights.

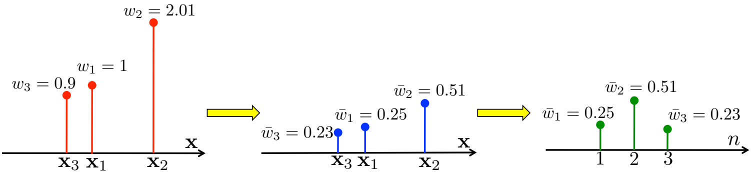
- Several methods (*particle filters, population Monte Carlo, adaptive importance sampling schemes*) use this formula above.
- It is possible to show that

$$1 \leq P_N^{(2)}(\bar{\mathbf{w}}) \leq N. \qquad (5)$$

See **[Kong92,Robert10,Liu01]**.

Weaknesses of $P_N^{(2)}(\bar{\mathbf{w}})$

- Due to the several approximations and strong assumptions: **loss of information**



- $P_N^{(2)}$ does not depend on $h(\mathbf{x})$.
- $P_N^{(2)}$ does not depend on the samples $\mathbf{x}_n$.
- One assumption is that $\mathbf{x}_n \sim q(\mathbf{x})$ for all n, but in different methods, we have $\mathbf{x}_1 \sim q_1(\mathbf{x}), \ldots, \mathbf{x}_N \sim q_N(\mathbf{x})$.
- By definition of $ESS = N\frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}$, we can have

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

Strengths/Reasons to use $P_N^{(2)}(\bar{\mathbf{w}})$

- Why is it used? It works reasonable well in different applications: people like it.
- Using only the info of the normalized weights $\bar{w}_n$, the inequalities $1 \leq P_N^{(2)}(\bar{\mathbf{w}}) \leq N$ are reasonable. It applies an *optimistic approach*:

$$\bar{\mathbf{w}}^* = \left[\frac{1}{N}, \ldots, \frac{1}{N}\right] \Longrightarrow P_N^{(2)}(\bar{\mathbf{w}}^*) = N, \tag{6}$$

$$\bar{\mathbf{w}}^{(i)} = [0, \ldots, \underbrace{1}_{i}, \ldots, 0] \Longrightarrow P_N^{(2)}(\bar{\mathbf{w}}^{(i)}) = 1. \tag{7}$$

- easy to be used, for adaptive resampling:

$$P_N^{(2)}(\bar{\mathbf{w}}) \leq \epsilon N$$

with $0 < \epsilon < 1$.

Alternative derivation/motivation of $P_N^{(2)}(\bar{\mathbf{w}})$

- Alternative derivation based on the need (or not) of resampling.
- Let us consider the Euclidean distance L_2 between these two the discrete uniform pmf $\mathcal{U}\{1, 2, \ldots, N\}$ and the pmf $\bar{w}_n$, i.e,

$$
\begin{aligned}
L_2 &= \sqrt{\sum_{n=1}^{N}\left(\bar{w}_n - \frac{1}{N}\right)^2} \\
&= \sqrt{\left(\sum_{n=1}^{N}\bar{w}_n^2\right) + N\left(\frac{1}{N^2}\right) - \frac{2}{N}\sum_{n=1}^{N}\bar{w}_n} \\
&= \sqrt{\left(\sum_{n=1}^{N}\bar{w}_n^2\right) - \frac{1}{N}} \\
&= \sqrt{\frac{1}{P_N^{(2)}(\bar{\mathbf{w}})} - \frac{1}{N}}. \qquad (8)
\end{aligned}
$$

- Maximizing $P_N^{(2)}$ corresponds to minimize L_2.

So far: Little summary

- So far: $\widehat{ESS} = P_N^{(2)}(\bar{\mathbf{w}})$ is a "bad" approximation of the theoretical definition.
- But the people like and use it; the main reason: maximizing $P_N^{(2)}$ corresponds to minimize L_2.
- Discrepancy/distance between pmf $\bar{\mathbf{w}}$ and uniform pmf $1/N$.
- Are there alternatives of the same type?

So far: Little summary

- So far: $\widehat{ESS} = P_N^{(2)}(\bar{\mathbf{w}})$ is a "bad" approximation of the theoretical definition.
- But the people like and use it; the main reason: maximizing $P_N^{(2)}$ corresponds to minimize L_2.
- Discrepancy/distance between pmf $\bar{\mathbf{w}}$ and uniform pmf $1/N$.
- Are there alternatives of the same type?
- (PS: the formula $P_N^{(2)}(\bar{\mathbf{w}})$ is also known as Kish's Effective Sample Size and is used in other branches of statistics that involve *weighted* samples)
- (PS2: for *correlated* samples, we have another formula)

Alternatives: ESS approx based on discrepancy

- Other authors also propose the *perplexity* measure based on the discrete entropy **[Cappe08]**.

Alternatives: ESS approx based on discrepancy

- Other authors also propose the *perplexity* measure based on the discrete entropy **[Cappe08]**.
- We can also consider

$$\widehat{ESS} = D_N^{(\infty)}(\bar{\mathbf{w}}) = \frac{1}{\max\left[\bar{w}_1, \ldots, \bar{w}_N\right]}, \tag{9}$$

- Note that $1 \leq D_N^{(\infty)}(\bar{\mathbf{w}}) \leq N$.

Generalized ESS functions

- Generalized ESS (G-ESS) function:

$$E_N(\bar{\mathbf{w}}) = E_N(\bar{w}_1, \ldots, \bar{w}_N) : \mathcal{S}_N \to [1, N], \tag{10}$$

where $\mathcal{S}_N \subset \mathbb{R}^N$ represents the *unit simplex*, namely,

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

Recall, we denote the vertices of the unit simplex as

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

and we denote also

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

Generalized ESS: *strictly required* conditions

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

$$E_N(\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_N) = E_N(\bar{w}_{j_1}, \bar{w}_{j_2}, \ldots, \bar{w}_{j_N}), \qquad (11)$$

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. (11) and has value N, i.e.,

$$E_N(\bar{\mathbf{w}}^*) = N \geq E_N(\bar{\mathbf{w}}). \qquad (12)$$

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),

$$E_N(\bar{\mathbf{w}}^{(j)}) = 1 \leq E_N(\bar{\mathbf{w}}). \qquad (13)$$

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

Generalized ESS: *welcome* conditions (1)

C4. **Unicity of extreme values:** The maximum at $\bar{\mathbf{w}}^*$ is unique, i.e., there are not other local maxima, and the the minimum value 1 is reached *only* at the vertices $\bar{\mathbf{w}}^{(j)}$, for all $j \in \{1, \ldots, N\}$.

GENERALIZED ESS: *welcome* CONDITIONS (2)

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

$$\bar{\mathbf{v}} = [\bar{v}_1, \dots, \bar{v}_{MN}] \in \mathbb{R}^{MN}, \quad M \geq 1, \tag{14}$$

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

$$\bar{\mathbf{v}} = \frac{1}{M}[\underbrace{\bar{\mathbf{w}}, \dots, \bar{\mathbf{w}}}_{M-times}], \tag{15}$$

i.e., $\bar{v}_1 = \frac{1}{M}\bar{w}_1, \dots, \bar{v}_N = \frac{1}{M}\bar{w}_N$ and $\bar{v}_{N+1} = \frac{1}{M}\bar{w}_1, \dots, \bar{v}_{MN} = \frac{1}{M}\bar{w}_N$. Then, the condition is given as

$$\frac{E_N(\bar{\mathbf{w}})}{N} = \frac{E_{MN}(\bar{\mathbf{v}})}{MN} \Longrightarrow E_N(\bar{\mathbf{w}}) = \frac{1}{M}E_{MN}(\bar{\mathbf{v}}), \tag{16}$$

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

Explanation of C5

- Following the *optimistic approach*, we would like, for instance,

$$\bar{\mathbf{w}} = \left[0, 0, \frac{1}{2}, \frac{1}{2}\right] \longrightarrow E_4(\bar{\mathbf{w}}) = 2,$$

and

$$\bar{\mathbf{v}} = \frac{1}{2}\left[\bar{\mathbf{w}}, \bar{\mathbf{w}}\right] = \left[0, 0, \frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{4}, \frac{1}{4}\right] \longrightarrow E_8(\bar{\mathbf{v}}) = 4.$$

i.e.,

$$E_4(\bar{\mathbf{w}}) = \frac{1}{2} E_8(\bar{\mathbf{v}}).$$

G-ESS: CLASSIFICATIONS

TABLE: 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

G-ESS: EXAMPLES (1)

- $P_N^{(2)}$ and $D_N^{(\infty)}$ are both of class PS, *proper and stable.*
- $V_N^{(0)}(\bar{\mathbf{w}}) = N - N_Z$; N_Z is the number of zeros, belongs to the class DS, *degenerate and stable.*
- Let us denote the harmonic mean of the normalized weights as

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

The following functions, involving the harmonic mean,

$$\begin{aligned} A_{1,N}(\bar{\mathbf{w}}) &= \frac{1}{(1-N)\mathrm{HarM}(\bar{\mathbf{w}})+1}, \\ A_{2,N}(\bar{\mathbf{w}}) &= (N^2-N)\mathrm{HarM}(\bar{\mathbf{w}})+1, \end{aligned}$$

are both *degenerate* G-ESS functions.

G-ESS: EXAMPLES (2)

They are *stable*:

- Perplexity **[Cappe08]**:

$$\mathrm{Per}_N(\bar{\mathbf{w}}) = 2^{H(\bar{\mathbf{w}})}, \quad \text{with} \quad H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log_2(\bar{w}_n).$$

- using Gini coefficient $G(\bar{\mathbf{w}})$:

$$\mathrm{Gin}_N(\bar{\mathbf{w}}) = -NG(\bar{\mathbf{w}}) + N.$$

- Threshold ESS (degenerate):

$$\text{N-Plus}_N(\bar{\mathbf{w}}) = N^+,$$

where $N^+ = \text{Cardinality}\{\bar{w}_n \geq \frac{1}{N}, \quad n = 1, \dots, N\}$.

Build G-ESS families

Given a non-linear transformation of the weights $f(\bar{\mathbf{w}})$

$f(\bar{\mathbf{w}})$: $\mathbb{R}^N \to \mathbb{R}$, 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 $\bar{\mathbf{w}}^* = \left[\frac{1}{N}, \ldots, \frac{1}{N}\right]$.
2. $f(\bar{\mathbf{w}})$ is symmetric in the sense of Eq. (11).
3. Considering the vertices of the unit simplex $\bar{\mathbf{w}}^{(i)} = \delta(i)$ in Eq. (B.2), then we also assume

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

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

We define the G-ESS families of type:

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

where we tune a and b in order to fulfill the strictly-required conditions (at least).

Build G-ESS families

We try to solve the linear ($f(\bar{\mathbf{w}}^*)$ and $f(\bar{\mathbf{w}}^{(i)})$ are given) systems

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

or

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

Four G-ESS families

Table: Summary of the G-ESS families (in general, *proper*, with exception...).

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

They satisfy always C1, C2, C3, often C4 (not always) and sometimes C5.

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

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

Special case $P_N^{(1)}(\bar{\mathbf{w}})$

- $a_r \to \pm\infty$, $b_r \to \mp\infty$, $(\bar{w}_n)^r \to 1$ when $r \to 1$, we have an indeterminate form of type $\frac{0}{0}$ in limit

$$\lim_{r\to 1} P_N^{(r)}(\bar{\mathbf{w}}) = \lim_{r\to 1} \frac{N^{(2-r)} - N}{(1-N)\sum_{n=1}^N (\bar{w}_n)^r + N^{(2-r)} - 1} = \frac{0}{0},$$

- Applying the L'Hôpital's rule,

$$\begin{aligned} P_N^{(1)}(\bar{\mathbf{w}}) &= \lim_{r\to 1} \frac{-N^{(2-r)}\log(N)}{-N^{(2-r)}\log(N) - (N-1)\sum_{n=1}^N \bar{w}_n^r \log(\bar{w}_n)}, \\ &= \frac{-N\log(N)}{-N\log(N) - (N-1)\sum_{n=1}^N \bar{w}_n \log(\bar{w}_n)}, \\ &= \frac{-N\frac{\log_2(N)}{\log_2 e}}{-N\frac{\log_2(N)}{\log_2 e} - (N-1)\sum_{n=1}^N \bar{w}_n \frac{\log_2(\bar{w}_n)}{\log_2 e}}, \\ &= \frac{-N\log_2(N)}{-N\log_2(N) + (N-1)H(\bar{\mathbf{w}})}, \end{aligned} \tag{17}$$

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 $\bar{w}_n$, $n = 1, \ldots, N$.

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

Parameter:	$\mathbf{r} \longrightarrow \mathbf{0}$	$\mathbf{r} \longrightarrow \mathbf{1}$	$\mathbf{r} \longrightarrow \infty$
$D_N^{(r)}(\bar{\mathbf{w}}) =$	$\frac{1}{(1-N)\mathrm{GeoM}(\bar{\mathbf{w}})+1}$	$\frac{-N\log_2(N)}{-N\log_2(N)+(N-1)H(\bar{\mathbf{w}})}$	$\frac{1}{\max[\bar{w}_1,\ldots,\bar{w}_N]}$
Comments:	*Geometric Mean* $\mathrm{GeoM}(\bar{\mathbf{w}}) = \left[\prod_{n=1}^N \bar{w}_n\right]^{1/N}$ *Degenerate*	*Discrete entropy* $H(\bar{\mathbf{w}}) = -\sum_{n=1}^N \bar{w}_n \log_2(\bar{w}_n)$ *Proper*	$D_N^{(\infty)}$ *Proper-Stable*

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

Parameter:	$\mathbf{r} \to \mathbf{0}$	$\mathbf{r} \to \mathbf{1}$	$\mathbf{r} \to \infty$
$V_N^{(r)}(\bar{\mathbf{w}}) =$	$N - N_Z$	$\frac{N-1}{\log_2(N)} H(\bar{\mathbf{w}}) + 1$	$\begin{cases} N & \text{if } \bar{\mathbf{w}} \neq \bar{\mathbf{w}}^{(i)}, \\ 1, & \text{if } \bar{\mathbf{w}} = \bar{\mathbf{w}}^{(i)}. \end{cases}$
Comments:	N_Z number of zeros in $\bar{\mathbf{w}}$ *Degenerate-Stable*	*Discrete Entropy* $H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log_2(\bar{w}_n)$ *Proper*	*Degenerate*

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

Par.:	$\mathbf{r \to 0}$	$\mathbf{r = \frac{1}{2}}$	$\mathbf{r \to 1}$	$\mathbf{r \to \infty}$
$S_N^{(r)}(\bar{\mathbf{w}})$	$(N^2 - N)\mathrm{GeoM}(\bar{\mathbf{w}}) + 1$	$\left(\sum_{n=1}^N \sqrt{w_n}\right)^2$	$\frac{N-1}{\log_2(N)} H(\bar{\mathbf{w}}) + 1$	$N + 1 - N \max[\bar{w}_1, \ldots, \bar{w}_N]$
Com.:	*Geometric Mean* $\mathrm{GeoM}(\bar{\mathbf{w}}) = \left[\prod_{n=1}^N \bar{w}_n\right]^{1/N}$ *Degenerate*	*Prop-Stable*	*Discrete Entropy* $H(\bar{\mathbf{w}})$ *Proper*	*Proper*

Summary

Table: **Stable G-ESS functions** and related inequalities.

Threshold	*GINI*	$D_N^{(\infty)}(\bar{\mathbf{w}})$	$P_N^{(2)}(\bar{\mathbf{w}})$	*Perplex*	$S_N^{(\frac{1}{2})}(\bar{\mathbf{w}})$	$V_N^{(0)}(\bar{\mathbf{w}})$
N^+	$-NG(\bar{\mathbf{w}}) + N$	$\frac{1}{\max[\bar{w}_1,\dots,\bar{w}_N]}$	$\frac{1}{\sum_{n=1}^N \bar{w}_n^2}$	$2^{H(\bar{\mathbf{w}})}$	$\left(\sum_{n=1}^N \sqrt{\bar{w}_n}\right)^2$	$N - N_Z$
DS all-C4	PS all	PS all	PS all	PS all	PS all	DS all-C4

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

Huggins-Roy's family

All proper and stable!! for $\beta > 0$ ($\beta = 0$ degenerate)

The Huggins-Roy's family introduced in [13] is defined as

$$\begin{aligned} H_N^{(\beta)}(\bar{\mathbf{w}}) &= \left(\frac{1}{\sum_{n=1}^{N} \bar{w}_n^{\beta}} \right)^{\frac{1}{\beta-1}}, \\ &= \left(\sum_{n=1}^{N} \bar{w}_n^{\beta} \right)^{\frac{1}{1-\beta}}, \qquad \beta \geq 0. \end{aligned}$$

Table 1 *Special cases of G-ESS functions contained in the Huggins-Roy's family.*

$\beta = 0$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$	$\beta = \infty$
$N - N_Z$ *where N_Z is the number of zeros in $\bar{\mathbf{w}}$*	$\left(\sum_{n=1}^{N} \sqrt{\bar{w}_n}\right)^2$	$\exp\left(-\sum_{n=}^{N} \bar{w}_n \log \bar{w}_n\right)$ *(perplexity)*	$\frac{1}{\sum_{n=1}^{N} \bar{w}_n^2}$ *(standard approximation)*	$\frac{1}{\max[\bar{w}_1, \ldots, \bar{w}_N]}$

Relationship with the Rényi entropy

The Rényi entropy [6] is defined as

$$R_N^{(\beta)}(\bar{\mathbf{w}}) = \frac{1}{1-\beta} \log \left[\sum_{n=}^{N} \bar{w}_n^{\beta} \right], \qquad \beta \geq 0,$$

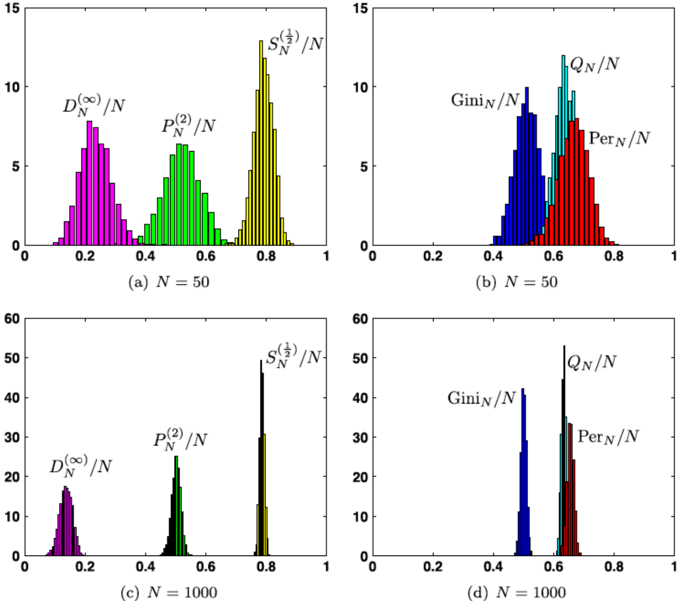
Then, it is straightforward to note that

$$H_N^{(\beta)}(\bar{\mathbf{w}}) = \exp\left(R_N^{(\beta)}(\bar{\mathbf{w}}) \right),$$

The Huggins-Roy's family contains the *diversity indices* based on the Rényi entropy.

Numerical Simulations/Considerations

Drawing $\bar{\mathbf{w}}$ uniformly in the unit simplex:



Numerical Simulations/Considerations

- These values (the statistics of these histograms) can be useful in the adaptive resampling applications:

$$\widehat{ESS}(\bar{\mathbf{w}}) \le \epsilon N,$$

with $0 < \epsilon < 1$.

- Perhaps, these histograms explain the value $\epsilon = \frac{1}{2}$, suggested in **[Doucet08; page15]** for $P_N^{(2)}$.
- We compare $P_N^{(2)}$ and $D_N^{(\infty)}$ in a "right way" within a particle filter (in that example $D_N^{(\infty)}$ works better)

$$\bar{\pi}(x) = \mathcal{N}(x; 0, 1), \tag{39}$$

and also a Gaussian proposal pdf,

$$q(x) = \mathcal{N}(x; \mu_p, \sigma_p^2), \tag{40}$$

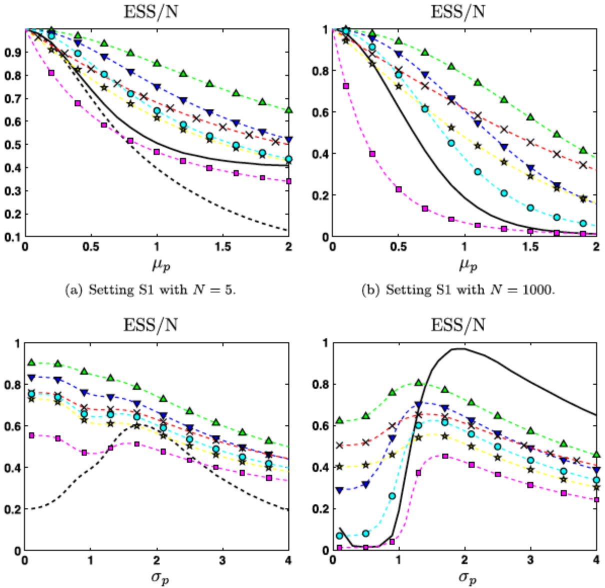
with mean μ_p and variance $\sigma_p{}^2$. Furthermore, we consider different experiment settings:

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, $q(x) \equiv \bar{\pi}(x)$. As μ_p increases, the proposal becomes more different from $\bar{\pi}$. We consider the estimation of the expected value of the random variable $X \sim \bar{\pi}(x)$, 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 $h(x) = x$.

Numerical Simulations

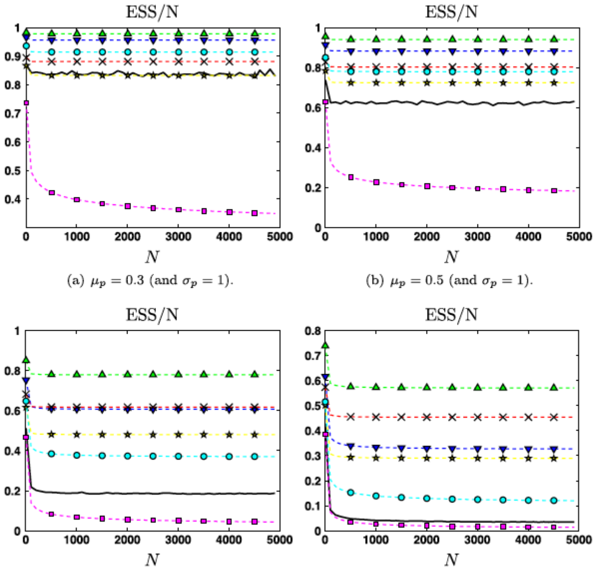


(a) Setting S1 with $N = 5$. (b) Setting S1 with $N = 1000$.

(c) Setting S2 with $N = 5$. (d) Setting S2 with $N = 1000$.

Fig. 3. ESS rates corresponding to $ESS_{var}(h)$ (solid line), $ESS_{MSE}(h)$ (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).

Numerical Simulations



(a) $\mu_p = 0.3$ (and $\sigma_p = 1$). (b) $\mu_p = 0.5$ (and $\sigma_p = 1$).

(c) $\mu_p = 1$ (and $\sigma_p = 1$). (d) $\mu_p = 1.5$ (and $\sigma_p = 1$).

Fig. 4. [Setting S3] ESS rates as function of N, corresponding to the theoretical ESS, i.e., $ESS_{var}(h)$ (solid line), and the G-ESS functions: $P_N^{(2)}$ (circles), $D_N^{(\infty)}$ (squares), $Gini_N$ (stars), $S_N^{(1/2)}$ (triangles up), Q_N (x-marks), Per_N (triangles down).

Numerical Simulations

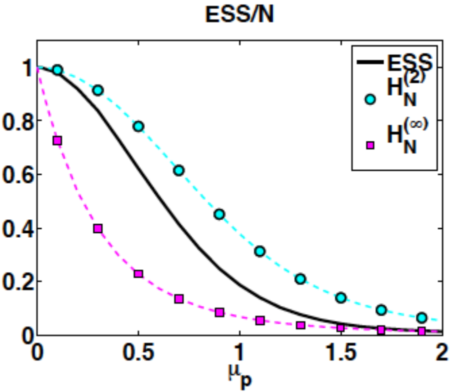


Figure 3 *ESS rates (i.e., the ratio of ESS values over N) corresponding to the theoretical ESS value (solid line), $H_N^{(2)}$ (circles) and $H_N^{(\infty)}$ (squares). We set* $N = 1000$.

Numerical Simulations

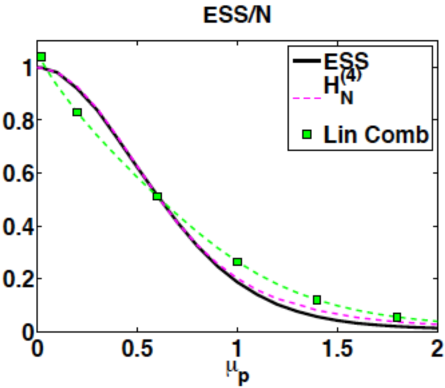


Figure 4 *ESS rates (i.e., the ratio of ESS values over N) corresponding to the theoretical ESS value (solid line), $H_N^{(4)}$ (dashed line) and the linear combination E_N in Eq. (5.4)-(5.5) (squares). We set $N = 1000$. The approximation provided by $H_N^{(4)}$ is virtually perfect for $\mu_p \leq 1$.*

Some references

[Kong92]: A. Kong. **A note on importance sampling using standardized weights**. *University of Chicago, Dept. of Statistics, Tech. Rep.*, vol. 348, 1992.

[Robert10]: C. P. Robert and G. Casella. **Introducing Monte Carlo Methods with R**. *Springer*, 2010.

[Liu01]: J. S. Liu, R. Chen, and T. Logvinenko. **A Theoretical Framework for Sequential Importance Sampling with Resampling**. *CHAPTER 11: Sequential Monte Carlo Methods in Practice, Springer, New York*, 2001.

[Cappe08]: O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert. **Population Monte Carlo**. *Statistics and Computing*, 18:447-459, 2008.

[Cappe04]: 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.

[Doucet01]: A. Doucet, N. de Freitas, and N. Gordon, editors. **Sequential Monte Carlo Methods in Practice**. *Springer, New York*, 2001.

[Doucet08]: A. Doucet, A. M. Johansen. **A Tutorial on Particle Filtering and Smoothing: Fifteen years later**. *Tec. Rep.*, 2008.