

Chapter 4

Iterated Jackknives and Two-Sided Variance Inequalities



Olivier Bousquet and Christian Houdré

Abstract We consider the variance of a function of n independent random variables and provide inequalities that generalize previous results obtained for i.i.d. random variables. In particular we obtain upper and lower bounds on the variance based on iterated jackknife statistics that can be considered generalizations of the Efron–Stein inequality.

4.1 Introduction

The properties of functions of n independent random variables, and in particular the estimation of their moments from the moments of their increments (i.e. when replacing a random variable by an independent copy) have been thoroughly studied (see, e.g., [2] for a comprehensive overview). We focus here on the variance and consider how to refine and generalize known extensions of the Efron–Stein inequality in the non-symmetric, non-iid case.

But first, let us review some of the existing results. Let X_1, X_2, \dots, X_n be iid random variables and let $S : \mathbb{R}^n \rightarrow \mathbb{R}$ be a statistic of interest which is symmetric, i.e., invariant under any permutation of its arguments, and square integrable. The (original) Efron–Stein inequality [3], states that the jackknife estimates of variance is biased upwards, i.e., denoting by \tilde{X} an independent copy of X_1, \dots, X_n , and setting $S_i = S(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n, \tilde{X})$, $i = 1, \dots, n$, and $S_{n+1} = S$, then

$$\text{Var } S \leq \mathbb{E}J_1, \quad (4.1)$$

O. Bousquet
Google Research, Brain Team, Zürich, Switzerland

C. Houdré (✉)
Georgia Institute of Technology, School of Mathematics, Atlanta, GA, USA
e-mail: houdre@math.gatech.edu

where

$$J_1 = \sum_{i=1}^{n+1} (S_i - \bar{S})^2 = \frac{1}{(n+1)} \sum_{1 \leq i < j \leq n+1} (S_i - S_j)^2, \quad (4.2)$$

and $\bar{S} = \sum_{i=1}^{n+1} S_i / (n+1)$. Beyond the original framework, the inequality (4.1) has seen many extensions and generalizations with different proofs which are well described in [2], whose notation we essentially adopt and to which we refer for a more complete bibliography and many instances of applications. Let us just say that (4.1) can be seen as the “well known” tensorization property of the variance which asserts that if X_1, X_2, \dots, X_n are independent random variables with $X_i \sim \mu_i$, then

$$\text{Var}_{\mu^n} S \leq \mathbb{E}_{\mu^n} \sum_{i=1}^n \text{Var}_{\mu_i} S, \quad (4.3)$$

where \mathbb{E}_{μ^n} and Var_{μ^n} are respectively the expectation and variance with respect to μ^n , the joint law of X_1, X_2, \dots, X_n , while $\text{Var}_{\mu_i} S$ is the variance of S with respect to μ_i , the law of X_i . In fact, if for each $i = 1, 2, \dots, n$, $\tilde{X}_i \sim \tilde{\mu}_i$ is an independent copy of X_i , then (4.3) can be rewritten as

$$\begin{aligned} \text{Var}_{\mu^n} S &\leq \frac{1}{2} \mathbb{E}_{\mu^n} \sum_{i=1}^n \mathbb{E}_{\mu_i \otimes \tilde{\mu}_i} (S - S_i)^2 \\ &= \frac{1}{2} \mathbb{E}_{\mu^n} \sum_{i=1}^n \mathbb{E}_{\tilde{\mu}_i} (S - S_i)^2, \end{aligned} \quad (4.4)$$

where $S_i = S(X_1, \dots, X_{i-1}, \tilde{X}_i, X_{i+1}, \dots, X_n)$.

Neither (4.1) nor (4.4), whose proof can be obtained, for example, by induction, require S to be symmetric. In case S is symmetric, and the random variables are identically distributed, the right-hand side of (4.4) becomes $n \mathbb{E}_{\mu^n \otimes \tilde{\mu}_1} (S - S_1)^2 / 2$ while, via (4.2), the right-hand side of (4.1) becomes $\binom{n-1}{2} \mathbb{E} (S_1 - S_2)^2 / (n+1) = n \mathbb{E} (S_1 - S_2)^2 / 2$, and (4.4) and (4.1) are identical.

Since the jackknife estimate of variance is biased upwards, it is natural to try to estimate the bias $\mathbb{E}J_1 - \text{Var} S$, and such an attempt is already presented in [5] via the “iterated jackknives”. Let us recall what was meant there: Resampling the jackknife statistics, introduce for any $k = 2, \dots, n$, the iterated jackknives J_2, J_3, \dots, J_n , leading to both upper and lower bounds on $\text{Var} S$, showing, in particular, that

$$\frac{1}{2} \mathbb{E}J_2 - \frac{1}{6} \mathbb{E}J_3 \leq \mathbb{E}J_1 - \text{Var} S \leq \frac{1}{2} \mathbb{E}J_2. \quad (4.5)$$

In [5], the inequalities (4.1) and (4.5) were viewed as statistical versions of generalized (multivariate) Gaussian Poincaré inequalities previously obtained in [6]. Indeed, setting $\nabla S := (S - S_1, S - S_2, \dots, S - S_n)$, then $\mathbb{E}J_1 = \mathbb{E}\|\nabla S\|^2$. If instead of looking at the vector of first differences, one looks at second and third ones, then the corresponding norms will lead to (4.5). Throughout the years, it was asked whether or not an inequality such as (4.5) would have a general version and a positive answer had been informally given. The aim of the present note is to provide a synthetic proof of these, removing the iid and symmetry assumptions in (4.5) and its generalizations, leading to generic inequalities. This could be useful, as these dormant inequalities seem to have found, in recent times, some new life, e.g., see [1, 8, 9].

4.2 Iterated Jackknife Bounds

Throughout and unless otherwise noted, X_1, \dots, X_n are independent random variables and $S : \mathbb{R}^n \rightarrow \mathbb{R}$ is a Borel function such that $\mathbb{E}S^2(X_1, \dots, X_n) < +\infty$. Next, and if S is short for $S(X_1, \dots, X_n)$, let, for any $i = 1, \dots, n$, $\mathbb{E}^{(i)}$ denote the conditional expectation with respect to the σ -field generated by $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$. Hence,

$$\begin{aligned} \mathbb{E}^{(i)} S &:= \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \\ &= \int_{-\infty}^{+\infty} S(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n) \mu_i(dx_i), \end{aligned} \quad (4.6)$$

where μ_i is the law of X_i . By convention, $\mathbb{E}^{(0)}$ is the identity operator and so $\mathbb{E}^{(0)} S = S$. Iterating the above, it is clear that

$$\begin{aligned} \mathbb{E}^{(i)} \mathbb{E}^{(j)} S &= \mathbb{E}^{(j)} \mathbb{E}^{(i)} S = \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n) \\ &:= \mathbb{E}^{(i,j)} S = \mathbb{E}^{(j,i)} S, \end{aligned} \quad (4.7)$$

for any $i, j = 1, \dots, n$ and that for $i = 0, 1, \dots, n$,

$$\mathbb{E}^{(i)} \mathbb{E}^{(0)} S = \mathbb{E}^{(0)} \mathbb{E}^{(i)} S := \mathbb{E}^{(i,0)} S = \mathbb{E}^{(0,i)} S = \mathbb{E}^{(i)} S.$$

Next, let

$$\text{Var}^{(i)} S := \mathbb{E}^{(i)} (S - \mathbb{E}^{(i)} S)^2 = \mathbb{E}^{(i)} S^2 - (\mathbb{E}^{(i)} S)^2,$$

$i = 0, 1, \dots, n$, and for any $i, j = 0, 1, \dots, n$, set

$$\text{Var}^{(i,j)} S := \mathbb{E}^{(i)} \text{Var}^{(j)} S - \text{Var}^{(j)} \mathbb{E}^{(i)} S = \text{Var}^{(j,i)} S \geq 0. \quad (4.8)$$

where, above, the rightmost equality follows from the commutativity property of the conditional expectations, as given in (4.7), while the inequality follows from convexity, and more precisely from the conditional Hölder's inequality.

Continuing with our notation, for any $i = 1, \dots, n$, let throughout \mathbb{E}_i denote the conditional expectation with respect to the σ -field generated by X_1, \dots, X_i , i.e., $\mathbb{E}_i S := \mathbb{E}(S \mid X_1, \dots, X_i)$, while this time $\mathbb{E}_0 S = \mathbb{E}S$.

At this point we also note that although $\text{Var}^{(i)}$ is the conditional variance with respect to the σ -field generated by $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, $\text{Var}^{(i,j)}$ is not the conditional variance with respect to the σ -field generated by $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n$. Indeed,

$$\text{Var}^{(i,j)} S = \mathbb{E}^{(i,j)}(S - \mathbb{E}^{(i,j)} S)^2 - \text{Var}^{(i)} \mathbb{E}^{(j)} S - \text{Var}^{(j)} \mathbb{E}^{(i)} S. \quad (4.9)$$

Further iterating, for $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$, then $\mathbb{E}^{(i_1)} \dots \mathbb{E}^{(i_k)} := \mathbb{E}^{(i_1, i_2, \dots, i_k)}$ is uniquely defined, i.e., the order in which the indices are taken is irrelevant, in particular $\mathbb{E}^{(1, 2, \dots, n)} S = \mathbb{E}S$. Still, iterating, set

$$\text{Var}^{(i_1, i_2, \dots, i_k)} S := \mathbb{E}^{(i_1)} \text{Var}^{(i_2, \dots, i_k)} S - \text{Var}^{(i_2, \dots, i_k)} \mathbb{E}^{(i_1)} S, \quad (4.10)$$

where again, above, the order in which the indices $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$ are taken is irrelevant, and further, by convexity, (4.10) is non-negative, i.e.,

$$\text{Var}^{(i_1, i_2, \dots, i_k)} S \geq 0.$$

With the help of the above definitions, and in view of [5], let us now introduce the iterated jackknives,

$$J_k := \sum_{1 \leq i_1 \neq i_2 \neq \dots \neq i_k \leq n} \text{Var}^{(i_1, \dots, i_k)} S = k! \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \text{Var}^{(i_1, \dots, i_k)} S.$$

Clearly, $J_1 = \sum_{i=1}^n \text{Var}^{(i)} S$ and in view of (4.6), (4.3) can just be rewritten as:

$$\text{Var} S \leq \mathbb{E} \sum_{i=1}^n \text{Var}^{(i)} S = \mathbb{E} J_1. \quad (4.11)$$

Still in view of the results of [5], we now intend to prove:

Theorem 4.2.1 *For any $p = 1, 2, \dots, [n/2]$,*

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k \leq \text{Var} S \leq \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k, \quad (4.12)$$

and

$$\text{Var } S = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} \mathbb{E} J_k. \quad (4.13)$$

Proof The proof of (4.13) is a simple decomposition/induction, while that of (4.12) further uses convexity. For $k = 1, 2, \dots, n$, let

$$R_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} \text{Var}^{(i_1, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1-1)} S),$$

with the understanding that for $i = 1$, $\mathbb{E}^{(1, i-1)} S = \mathbb{E}^{(0)} S = S$. Then, first note that,

$$\begin{aligned} \mathbb{E} R_1 &= \mathbb{E} \sum_{i_1=1}^n \left((\mathbb{E}^{(1, \dots, i_1-1)} S)^2 - (\mathbb{E}^{(1, \dots, i_1)} S)^2 \right) \\ &= \mathbb{E} (S^2 - (\mathbb{E} S)^2) = \text{Var} S. \end{aligned} \quad (4.14)$$

Notice further that for $2 \leq k \leq n-1$,

$$\begin{aligned} \mathbb{E} R_k &= \mathbb{E} \sum_{1 \leq i_1 < \dots < i_k \leq n} \text{Var}^{(i_1, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1-1)} S) \\ &= \mathbb{E} \sum_{1 \leq i_1 < \dots < i_k \leq n} \left(\text{Var}^{(i_2, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1-1)} S) - \text{Var}^{(i_2, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1)} S) \right) \\ &= \mathbb{E} \sum_{1 < i_2 < \dots < i_k \leq n} \sum_{i_1=1}^{i_2-1} \left(\text{Var}^{(i_2, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1-1)} S) - \text{Var}^{(i_2, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1)} S) \right) \\ &= \mathbb{E} \sum_{1 \leq i_2 < \dots < i_k \leq n} \left(\text{Var}^{(i_2, \dots, i_k)} S - \text{Var}^{(i_2, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_2-1)} S) \right) \\ &= \frac{\mathbb{E} J_{k-1}}{(k-1)!} - \mathbb{E} R_{k-1}. \end{aligned} \quad (4.15)$$

Finally, it is clear that, $R_n = \text{Var}^{(1, \dots, n)} S$, and so $n! \mathbb{E} R_n = \mathbb{E} J_n$. Combining the last three identities, gives (4.13). To obtain (4.12), note first that by convexity and for any $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

$$\mathbb{E}^{(1, \dots, i_1-1)} \text{Var}^{(i_1, \dots, i_k)} S \geq \text{Var}^{(i_1, \dots, i_k)} (\mathbb{E}^{(1, \dots, i_1-1)} S). \quad (4.16)$$

Hence, taking expectation and summing gives $\mathbb{E} J_k \geq k! \mathbb{E} R_k$, which when combined with (4.15) finishes the proof. \square

Remark 4.2.2

- (i) In case S is symmetric, i.e., invariant under any permutation of its arguments, $J_k = n(n-1) \dots (n-k+1) \text{Var}^{(1, \dots, k)} S$, then $\mathbb{E}J_k = n(n-1) \dots (n-k+1) \mathbb{E} \text{Var}^{(1, \dots, k)} S$, and (4.13) and (4.12) precisely recover corresponding results in [5].
- (ii) The inequalities (4.12) can be viewed as martingale inequalities.
- (iii) As in [2] or [1], one could also rewrite (4.12) using only the positive or negative parts of the involved quantities.
- (iv) It is natural to wonder whether or not the above inequalities have Φ -entropic versions; this will be explored and presented elsewhere.

Let us now further refine (4.12) providing, in particular, a non-trivial lower bound on the bias $\mathbb{E}J_1 - \text{Var } S$ improving upon (4.5). To do so, denote by $(\bar{i}_1, \dots, \bar{i}_k)$ the complement of the indices (i_1, \dots, i_k) (i.e., the ordered sequence of elements of the set $\{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$), and introduce the following quantities:

$$K_k := k! \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \text{Var}^{(i_1, \dots, i_k)} \overline{\mathbb{E}^{(i_1, \dots, i_k)} S}.$$

It is clear that by Jensen's inequality and the convexity of $\text{Var}^{(i_1, \dots, i_k)}$ we have

$$\mathbb{E}K_k \leq \mathbb{E}J_k.$$

Theorem 4.2.3 For any $p = 1, 2, \dots, \lfloor n/2 \rfloor$,

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k + \frac{1}{(2p+1)!} \mathbb{E}K_{2p+1} \leq \text{Var } S \leq \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k - \frac{1}{(2p)!} \mathbb{E}K_{2p} \quad (4.17)$$

Proof The only modification compared to the proof of Theorem 4.2.1 is that instead of using the bound $\mathbb{E}J_k \geq k! \mathbb{E}R_k$ we use the fact that

$$\mathbb{E}K_k \leq k! \mathbb{E}R_k,$$

which follows from the convexity of $\text{Var}^{(i_1, \dots, i_k)}$. □

In particular, from Theorems 4.2.1 and 4.2.3, the following inequalities hold true:

$$\frac{1}{2} \mathbb{E}K_2 \leq \mathbb{E}J_1 - \text{Var } S \leq \frac{1}{2} \mathbb{E}J_2. \quad (4.18)$$

4.3 Relationship with the Hoeffding Decomposition

Let us recall the notion of a Hoeffding decomposition [4] (see [7] Section 2 for the general non-symmetric non-iid case). Given a function $f(X) \in L_1(\mathbb{P})$, it is the unique decomposition

$$\begin{aligned} f(X_1, \dots, X_n) &= \mathbb{E}f(X) + \sum_{1 \leq i \leq n} h_i(X_i) + \sum_{1 \leq i < j \leq n} h_{ij}(X_i, X_j) + \dots \\ &= f_0 + f_1 + \dots + f_n, \end{aligned}$$

such that $\mathbb{E}^{(i_s)} h_{i_1, \dots, i_k}(X_{i_1}, \dots, X_{i_k}) = 0$ whenever $1 \leq i_1 < \dots < i_k \leq n$, $s = 1, \dots, k$. The term f_d is called the Hoeffding term of degree d and these terms form an orthogonal decomposition of f in $L_2(\mathbb{P})$ (provided, of course, $f \in L_2(\mathbb{P})$), so that $\text{Var} f = \sum_{k=1}^n \text{Var} f_k = \sum_{I \subset \{1, \dots, n\}} \mathbb{E} h_I^2$

The following lemma provides a relationship between the previously introduced iterated jackknives and the variance of the Hoeffding terms.

Lemma 4.3.1 *For any k such that $1 \leq k \leq n$,*

$$\frac{1}{k!} \mathbb{E} J_k(f) = \sum_{j \geq k} \binom{j}{k} \text{Var} f_j,$$

and

$$\frac{1}{k!} \mathbb{E} K_k(f) = \text{Var} f_k.$$

Proof Let us rewrite the Hoeffding decomposition of f as $f = \sum_{I \subset \{1, 2, \dots, n\}} h_I$. We have $\mathbb{E}^{(i)} h_I = 0$ whenever $i \in I$, and $\mathbb{E}^{(i)} h_I = h_I$ otherwise. Hence, $\text{Var}^{(i)} h_I = \mathbb{E}^{(i)} h_I^2$ if $i \in I$ and 0 otherwise. Therefore, $\mathbb{E} \text{Var}^{(i)} S = \sum_{i \in I} \mathbb{E} h_I^2$.

Continuing with the same reasoning, we can see that $\text{Var}^{(i)} \mathbb{E}^{(j)} h_I = \mathbb{E}^{(i)} h_I^2$, if $i \in I$ and $j \notin I$ and 0 otherwise, thus $\mathbb{E} \text{Var}^{(i)} \mathbb{E}^{(j)} S = \sum_{i \in I, j \notin I} \mathbb{E} h_I^2$ so that $\mathbb{E} \text{Var}^{(i, j)} S = \sum_{\{i, j\} \subset I} \mathbb{E} h_I^2$ and by induction, we get that

$$\mathbb{E} \text{Var}^{(i_1, \dots, i_k)} S = \sum_{\{i_1, \dots, i_k\} \subset I} \mathbb{E} h_I^2.$$

If we now sum over the possible sets of indices, since each term $\mathbb{E} h_I^2$ appears as many times as there are subsets of size k of I , this implies that $\mathbb{E} J_k = k! \sum_{|I| \geq k} \binom{|I|}{k} \mathbb{E} h_I^2 = k! \sum_{j \geq k} \binom{j}{k} \text{Var} f_j$ and gives the first statement.

To prove the second statement of the lemma, observe that $\mathbb{E}^{\overline{(i_1, \dots, i_k)}} S = \sum_{I \subset \{i_1, \dots, i_k\}} h_I$ so that $\mathbb{E} \text{Var}^{(i_1, \dots, i_k)} \mathbb{E}^{\overline{(i_1, \dots, i_k)}} S = \mathbb{E} h_{i_1, \dots, i_k}^2$, and therefore $\mathbb{E} K_k = k! \sum_{|I|=k} \mathbb{E} h_I^2 = k! \text{Var} f_k$. \square

It is easily verified that (4.13) can be recovered as a consequence of Lemma 4.3.1. Also, from Lemma 4.3.1 it is easy to get the following corollary obtained in [1] (as part of their Theorem 1.8).

Corollary 4.3.2 *Let S have Hoeffding decomposition of type $S = \mathbb{E} S + \sum_{k=d}^n S_k$, i.e., such that $f_k = 0$, for $1 \leq k < d$, then*

$$\text{Var} S \leq \frac{1}{d!} \mathbb{E} J_d. \quad (4.19)$$

Proof Using the fact that $f_k = 0$, for $1 \leq k < d$, we have

$$\text{Var} S = \sum_{j=d}^n \text{Var} f_j \leq \sum_{j=d}^n \binom{j}{d} \text{Var} f_j = \frac{1}{d!} \mathbb{E} J_d,$$

where the last equality follows from Lemma 4.3.1. \square

Acknowledgements C. Houdré was partially supported by the grant # 524678 from the Simons Foundation and by a Simons Fellowship grant # 267336. This material is also based in part upon material supported by NSF grant No. 1440140 while this author was in residence at MSRI in Berkeley, CA, during Fall, 2017.

References

1. S.G. Bobkov, F. Götze, H. Sambale, Higher order concentration of measure. *Commun. Contemp. Math.* **21**, 1850043 (2019). arXiv:1709.06838
2. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities. A nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
3. B. Efron, C. Stein, The jackknife estimate of variance. *Ann. Stat.* **9**, 586–596 (1981)
4. W. Hoeffding, A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325 (1948)
5. C. Houdré, The iterated jackknife estimate of variance. *Statist. Probab. Lett.* **35**, 197–201 (1997)
6. C. Houdré, A. Kagan, Variance inequalities for functions of Gaussian variables. *J. Theoret. Probab.* **8**, 23–30 (1995)
7. S. Karlin, Y. Rinott, Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. Stat.* **10**, 485–501 (1982)
8. A. Maurer, A Bernstein-type inequality for functions of bounded interaction. *Bernoulli* **25**, 1451–1471 (2019). arXiv: 1701.06191
9. A. Maurer, M. Pontil, Empirical bounds for functions with weak interactions, in *COLT 2018*. arXiv:1803.03934