

# Iterated Jackknives And Two-Sided Variance Inequalities

O. Bousquet\* C. Houdré<sup>†‡</sup>

January 1, 2020

## Abstract

We consider the variance of a function of  $n$  independent random variables and provide new inequalities which, in particular, extend previous results obtained for symmetric functions in the i.i.d. setting. For instance, we obtain various upper and lower variance bounds based on iterated jackknives statistics that can be considered as generalizations of the Efron-Stein inequality.

## 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.

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

---

\*Google Research, Brain Team, 8002 Zürich, Switzerland.

<sup>†</sup>Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia, 30332-0160.

<sup>‡</sup>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 work supported by the National Science Foundation under Grant No.1440140, while this author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall semester of 2017.

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

$$\text{Var } S \leq \mathbb{E}J_1, \quad (1.1)$$

where

$$J_1 = \sum_{i=1}^{n+1} (S_i - \bar{S})^2 = \frac{1}{(n+1)} \sum_{1 \leq i < j \leq n+1} (S_i - S_j)^2, \quad (1.2)$$

and  $\bar{S} = \sum_{i=1}^{n+1} S_i / (n+1)$ . Beyond the original framework, the inequality (1.1) 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 (1.1) 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

$$\text{Var}_{\mu^n} S \leq \mathbb{E}_{\mu^n} \sum_{i=1}^n \text{Var}_{\mu_i} S, \quad (1.3)$$

where  $\mathbb{E}_{\mu^n}$  and  $\text{Var}_{\mu^n}$  are respectively the expectation and variance with respect to  $\mu^n := \mu_1 \otimes \dots \otimes \mu_n$ , 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  $\mu_i$ , 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 (1.3) can be rewritten as

$$\begin{aligned} \text{Var}_{\mu^n} S &\leq \frac{1}{2} \mathbb{E}_{\mu^n} \sum_{i=1}^n \mathbb{E}_{\mu_i \otimes \tilde{\mu}_i} (S - S_i)^2 \\ &= \frac{1}{2} \mathbb{E}_{\mu^n} \sum_{i=1}^n \mathbb{E}_{\tilde{\mu}_i} (S - S_i)^2, \end{aligned} \quad (1.4)$$

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

Neither (1.1) nor (1.4), 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 (1.4) becomes  $n\mathbb{E}_{\mu^n \otimes \tilde{\mu}_1}(S - S_1)^2/2$  while, via (1.2), the right-hand side of (1.1) becomes  $\binom{n+1}{2}\mathbb{E}(S_1 - S_2)^2/(n+1) = n\mathbb{E}(S_1 - S_2)^2/2$ , and (1.4) and (1.1) are equal.

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

$$\frac{1}{2}\mathbb{E}J_2 - \frac{1}{6}\mathbb{E}J_3 \leq \mathbb{E}J_1 - \text{Var } S \leq \frac{1}{2}\mathbb{E}J_2. \quad (1.5)$$

In [5], the inequalities (1.1) and (1.5) 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 (1.5). Throughout the years, it was asked whether or not an inequality such as (1.5) 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 (1.5) 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 [8], [1], [9].

## 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  $\sigma$ -field generated by  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ . Hence,

$$\begin{aligned} \mathbb{E}^{(i)} S &:= \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \\ &= \int_{-\infty}^{+\infty} S(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n) \mu_i(dx_i), \end{aligned} \quad (2.1)$$

where  $\mu_i$  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

$$\begin{aligned}\mathbb{E}^{(i)}\mathbb{E}^{(j)}S &= \mathbb{E}^{(j)}\mathbb{E}^{(i)}S = \mathbb{E}(S \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n) \\ &:= \mathbb{E}^{(i,j)}S = \mathbb{E}^{(j,i)}S,\end{aligned}\tag{2.2}$$

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

$$\text{Var}^{(i,j)}S := \mathbb{E}^{(i)}\text{Var}^{(j)}S - \text{Var}^{(j)}\mathbb{E}^{(i)}S = \text{Var}^{(j,i)}S \geq 0.\tag{2.3}$$

where, above, the rightmost equality follows from the commutativity property of the conditional expectations, as given in (2.2), while the inequality follows from convexity, and more precisely from the conditional Jensen's inequality.

At this point we also note that although  $\text{Var}^{(i)}$  is the conditional variance with respect to the  $\sigma$ -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  $\sigma$ -field generated by  $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n$ . Indeed,

$$\text{Var}^{(i,j)}S = \mathbb{E}^{(i,j)}(S - \mathbb{E}^{(i,j)}S)^2 - \text{Var}^{(i)}\mathbb{E}^{(j)}S - \text{Var}^{(j)}\mathbb{E}^{(i)}S.\tag{2.4}$$

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

$$\text{Var}^{(i_1, i_2, \dots, i_k)}S := \mathbb{E}^{(i_1)}\text{Var}^{(i_2, \dots, i_k)}S - \text{Var}^{(i_2, \dots, i_k)}\mathbb{E}^{(i_1)}S,\tag{2.5}$$

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, (2.5) is non-negative, i.e.,

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

Another set of identities, more in line with [5], is also easily obtained via iterated differences, namely,  $\mathbb{E}\text{Var}^{(i)}S = \mathbb{E}(S - S_i)^2/2$ , and iterating,

$$\mathbb{E}\text{Var}^{(i_1, i_2, \dots, i_k)}S = \frac{1}{2^k} \mathbb{E} \left( (S - S_{i_1})_{i_2, \dots, i_k} \right)^2. \quad (2.6)$$

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 (2.1), (1.3) can just be rewritten as:

$$\text{Var}S \leq \mathbb{E} \sum_{i=1}^n \text{Var}^{(i)}S = \mathbb{E}J_1. \quad (2.7)$$

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

**Theorem 2.1.** *For any  $p = 1, 2, \dots, [n/2]$ ,*

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k \leq \text{Var}S \leq \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k, \quad (2.8)$$

and

$$\text{Var}S = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k. \quad (2.9)$$

*Proof.* The proof of (2.9) is a simple decomposition/induction, while that of (2.8) 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,

$$\begin{aligned} \mathbb{E}R_1 &= \mathbb{E} \sum_{i_1=1}^n ((\mathbb{E}^{(1, \dots, i_1-1)}S)^2 - (\mathbb{E}^{(1, \dots, i_1)}S)^2) \\ &= \mathbb{E}(S^2 - (\mathbb{E}S)^2) = \text{Var}S. \end{aligned} \quad (2.10)$$

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

$$\begin{aligned}
\mathbb{E}R_k &= \mathbb{E} \sum_{1 \leq i_1 < \dots < i_k \leq n} \text{Var}^{(i_1, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1-1)} S) \\
&= \mathbb{E} \sum_{1 \leq i_1 < \dots < i_k \leq n} \left( \text{Var}^{(i_2, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1-1)} S) - \text{Var}^{(i_2, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1)} S) \right) \\
&= \mathbb{E} \sum_{1 < i_2 < \dots < i_k \leq n} \sum_{i_1=1}^{i_2-1} \left( \text{Var}^{(i_2, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1-1)} S) - \text{Var}^{(i_2, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1)} S) \right) \\
&= \mathbb{E} \sum_{1 \leq i_2 < \dots < i_k \leq n} \left( \text{Var}^{(i_2, \dots, i_k)} S - \text{Var}^{(i_2, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_2-1)} S) \right) \\
&= \frac{\mathbb{E}J_{k-1}}{(k-1)!} - \mathbb{E}R_{k-1}. \tag{2.11}
\end{aligned}$$

Finally, it is clear that,  $R_n = \text{Var}^{(1, \dots, n)} S = J_n/n!$ , and so  $n! \mathbb{E}R_n = \mathbb{E}J_n$ . Combining the last three identities, gives (2.9). To obtain (2.8), note first that by convexity and for any  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ,

$$\mathbb{E}^{(1, \dots, i_1-1)} \text{Var}^{(i_1, \dots, i_k)} S \geq \text{Var}^{(i_1, \dots, i_k)}(\mathbb{E}^{(1, \dots, i_1-1)} S). \tag{2.12}$$

Hence, taking expectation and summing gives  $\mathbb{E}J_k \geq k! \mathbb{E}R_k$ , which when combined with (2.11) finishes the proof.  $\square$

**Remark 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 (2.9) and (2.8) precisely recover corresponding results in [5].

(ii) The inequalities (2.8) can be viewed as martingale inequalities.

(iii) As in [2] or [1], one could also rewrite (2.8) using only the positive or negative parts of the involved quantities.

(iv) It is natural to wonder whether or not the above inequalities have  $\Phi$ -entropic versions; this will be explored and presented elsewhere.

Let us now further refine (2.8) providing, in particular, a non-trivial non-negative lower bound on the bias  $\mathbb{E}J_1 - \text{Var} S$  improving upon (1.5). To do so, denote by  $\overline{(i_1, \dots, i_k)}$  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, for  $k \geq 1$ , 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 2.3.** For any  $p = 1, 2, \dots, [n/2]$ ,

$$\sum_{k=1}^{2p} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k + \frac{1}{(2p+1)!} \mathbb{E}K_{2p+1} \leq \text{Var } S \leq \sum_{k=1}^{2p-1} \frac{(-1)^{k+1}}{k!} \mathbb{E}J_k - \frac{1}{(2p)!} \mathbb{E}K_{2p}. \quad (2.13)$$

*Proof.* The only modification compared to the proof of Theorem 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  $\text{Var}^{(i_1, \dots, i_k)}$ , i.e., from (2.12).  $\square$

In particular, from Theorem 2.1 and Theorem 2.3, (the case  $p = 0$ , being clear) the following inequalities hold true:

$$\begin{aligned} 0 &\leq \mathbb{E}K_1 \leq \text{Var } S \leq \mathbb{E}J_1, \\ 0 &\leq \frac{1}{2} \mathbb{E}K_2 \leq \mathbb{E}J_1 - \text{Var } S \leq \frac{1}{2} \mathbb{E}J_2. \end{aligned}$$

### 3 Relationship With The Hoeffding Decomposition

Let us recall the notion of Hoeffding decomposition [4] (see [3] or [7, Section 2] for the general non-symmetric non-iid case). Given an integrable random variable  $f(X)$ , 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 (provided, of course, that  $f(X)$

is square integrable); and so  $\text{Var} f = \sum_{k=1}^n \text{Var} f_k = \sum_{I \subset \{1, \dots, n\}} \mathbb{E} h_I^2$ , where  $I \neq \emptyset$ .

Continuing with our notation, for any  $i = 1, \dots, n$ , let  $\mathbb{E}_i$  denote the conditional expectation with respect to the  $\sigma$ -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  $\mathbb{E}_0 S = \mathbb{E} S$ .

Then, it is easily seen that above,  $f_0 = \mathbb{E}_0 f$ ,  $h_i = \mathbb{E}_i f - \mathbb{E}_0 f$ ,  $i = 1, \dots, n$ ,  $h_{ij} = \mathbb{E}_{ij} f - \mathbb{E}_i f - \mathbb{E}_j f + \mathbb{E}_0 f$ ,  $1 \leq i < j \leq n$ , etc.

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

**Lemma 3.1.** *For any  $k$  such that  $1 \leq k \leq n$ ,*

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

,

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

and so

$$\mathbb{E} J_k(f) = \sum_{j=k}^n \frac{1}{(j-k)!} \mathbb{E} K_j(f).$$

*Proof.* Rewrite the Hoeffding decomposition of  $f$  as  $f = \mathbb{E} f + \sum_{I \subset \{1, \dots, n\}} h_I$ , where again  $I \neq \emptyset$ . Then,  $\mathbb{E}^{(i)} h_I = 0$ , 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 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  $\mathbb{E} J_k = k! \sum_{|I|=k} \binom{|I|}{k} \mathbb{E} h_I^2 = k! \sum_{j=k}^n \binom{j}{k} \text{Var} f_j$ , proving the first statement.

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



To obtain the third statement, just combined the previous two.  $\square$

It is easily verified that (2.9) can be recovered from Lemma 3.1 and that

$$\text{Var } S = \mathbb{E}J_1 - \sum_{k=2}^n \frac{k-1}{k!} \mathbb{E}K_k. \quad (3.1)$$

Moreover, still from Lemma 3.1,

$$\text{Var } S = \sum_{k=1}^n \frac{1}{k!} \mathbb{E}K_k. \quad (3.2)$$

Lemma 3.1 also easily imply the following corollary obtained in [1] (as part of their Theorem 1.8) which moreover can be complemented with the trivial lower bound  $\mathbb{E}K_d/k! \leq \text{Var } S$ .

**Corollary 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*

$$\text{Var } S \leq \frac{1}{d!} \mathbb{E}J_d. \quad (3.3)$$

*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 3.1.  $\square$

## References

- [1] S. G. Bobkov, F. Götze and H. Sambale. Higher order concentration of measure. *Commun. Contemp. Math.*, **21**, (2019), arXiv:1709.06838.
- [2] S. Boucheron, G. Lugosi and P. Massart. Concentration inequalities. A nonasymptotic theory of independence. Oxford University Press, Oxford, (2013).
- [3] B. Efron and C. Stein. The jackknife estimate of variance. *Ann. Stat.*, **9**, (1981), 586-596,.

- [4] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, **19**, (1948), 293-325.
- [5] C. Houdré. The iterated jackknife estimate of variance. *Statist. Probab. Lett.*, **35**, (1997), 197-201.
- [6] C. Houdré and A. Kagan. Variance inequalities for functions of Gaussian variables. *J. Theoret. Probab.*, **8**, (1995), 233-0.
- [7] S. Karlin, and Y. Rinott. Applications of ANOVA type decompositions for comparisons of conditional variance statistics including jackknife estimates. *Ann. of Stat.*, **10**, (1982), 485-501.
- [8] A. Maurer. A Bernstein-type inequality for functions of bounded interaction. *Bernoulli*, **25**, (2019), 1451–1471. arXiv: 1701.06191.
- [9] A. Maurer and M. Pontil. Empirical bounds for functions with weak interactions. COLT 2018. arXiv:1803.03934.