A Note on Jackknife Based Estimates of Sampling Distributions

C. Houdré*

Abstract

Tail estimates of statistics are given; they depend on jackknife estimates of variance of the statistics of interest.

1 Introduction

We wish in these simple notes to provide some further possible answers to the following general question: Given a statistic S of the independent random sample $X_1, \cdots, X_n$, how to find estimates of the sampling distribution of S? Of course, any potential answer to that question will have to depend on the properties of the statistic and of the random sample. Since we strive for generality, we certainly do not wish to presuppose any distributional assumptions on the sample, and thus work in a non–parametric setting. The lack of parametric assumption will be overcome by using jackknife type estimates of variance which will control our estimates of the sampling distribution. Among the many methods developed to answer our main question, Efron's bootstrap and its ancestor the Quenouille–Tuckey jackknife have enjoyed tremendous success (see [E1], [E2], [ET], [M], [PS], [F]). It is, however, our belief that there is still a need for a theoretical understanding of our question and a attempt at answering it is provided below. Indeed, the jackknife and bootstrap methodologies have mainly been used to estimate some parameters of interest; but not really to estimate the sampling distribution itself which is our main goal in these notes. Of course, bootstrap or jackknife confidence intervals also provide sampling distribution estimates, but more precisely to assess the accuracy of the parameter estimates than to analyze the sampling distribution itself.

A main motivation for the present work is [H], where an expansion of the variance of a symmetric statistic of iid samples is obtained via the iterated jackknife. This expansion was also used to give generalizations of the Efron–Stein [ES] inequality which asserts that the jackknife estimate of variance is biased upwards. In particular, it was highlighted there than the jackknife could be thought of as a statistical gradient. This same idea will be repeatedly used here and will help us in answering our introductory

*Southeast Applied Analysis Center, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, houdre@math.gatech.edu; research supported in part by the NSF Grant No. DMS–9803239. This work was done in part while the author greatly enjoyed the hospitality of CIMAT, Guanajuato, Gto, Mexico.

[0]Key words: Jackknife, Efron-Stein Inequality, Deviation Inequality, Bootstrap.

question. A second main motivation is a functional approach to deviation inequalities which has been popularized by Ledoux [L1], [L2]. In this approach, one starts with a functional inequality such as a Poincaré or a log–Sobolev inequality leading to moment generating function estimates and then to deviation (this is the so called Herbst technique). From a statistical point of view requiring that the statistics of interest satisfy functional inequalities is rather unnatural (e.g., they might not be differentiable). It is however possible to apply such a methodology to obtain estimates on the sampling distribution of a statistic of interest. The key ingredient in this approach is to replace the gradient by the jackknife leading to a resampling versions of functional inequalities between, say, the entropy and its jackknife estimate. With this analogy, the approach in obtaining our results is somehow the converse of the functional approach. Starting with a statistic, we find the proper jackknife inequality and this lead to tail inequalities. The techniques of proofs can thus be seen as jackknife modifications of techniques in [L1], [L2].

Let us describe a little bit the content of our paper. In the next section, we recall the notion of entropy, and state a known tensorization property. This allows us to obtain our first sampling distribution result which gives normal tails estimates. This first result recovers, the classical Azuma–Hoeffding martingale inequality and it is also compared to various known estimates. A second result which leads to Poissonian type tails inequalities is then obtained. It recovers, in particular, a classical inequality of Bennett on sums of independent random variables. The final section is devoted to a further understanding of our approach and to the Efron–Stein inequality.

2 Entropy and Sampling Distribution

While jackknife based variance inequalities have antecedents, such is not the case for the entropy. Recall that given a statistics $S \geq 0$, (and with the convention $0 \log 0 = 0$) the entropy of S is given by

$$\mathbf{Ent} S = \mathbf{E} S \log S - \mathbf{E} S \log \mathbf{E} S. \tag{2.1}$$

(by Jensen's inequality, $\mathbf{Ent} S \geq 0$). We start this section with a known tensorization lemma (see [L1], [L2]). An equivalent formulation of this result is already present in [L] (see [HT] for further references). The proof below is only given for the reader's convenience and for the sake of completeness.

Lemma 2.1 *Let $X_1, X_2, \cdots, X_n$, be independent random variables with $X_i \sim F_i$, and let the statistic $S : \mathbf{R}^n \longrightarrow \mathbf{R}^+$ have finite entropy. Then,*

$$\mathbf{Ent}_{F^n} S \leq \mathbf{E}_{F^n} \sum_{i=1}^{n} \mathbf{Ent}_{F_i} S, \tag{2.2}$$

where $F^n = F_1 \times \cdots \times F_n$.

Proof. The proof of (2.1) is done by induction. For $n = 1$, the inequality in (2.1) becomes equality and so the result is clear. Now assume the result for $n - 1$, and let

$F^{n-1} = F_1 \times \cdots \times F_{n-1}$. Then we have

$$\begin{aligned}
\mathbf{Ent}_{F^n} S &= \mathbf{E}_{F^n} S \log S - \mathbf{E}_{F^n} S \log \mathbf{E}_{F^n} S \\
&= \mathbf{E}_{F_n} \mathbf{Ent}_{F^{n-1}} S + \mathbf{E}_{F_n} (\mathbf{E}_{F^{n-1}} S \log \mathbf{E}_{F^{n-1}} S) - \mathbf{E}_{F^n} S \log \mathbf{E}_{F^n} S \\
&\leq \mathbf{E}_{F_n} \mathbf{E}_{F^{n-1}} \sum_{i=1}^{n-1} \mathbf{Ent}_{F_i} S + \mathbf{E}_{F_n} (\mathbf{E}_{F^{n-1}} S \log \mathbf{E}_{F^{n-1}} S) - \mathbf{E}_{F^n} S \log \mathbf{E}_{F^n} S \\
&\leq \mathbf{E}_{F^n} \sum_{i=1}^{n-1} \mathbf{Ent}_{F_i} S + \mathbf{E}_{F_n} (\mathbf{E}_{F^{n-1}} S \log \mathbf{E}_{F^{n-1}} S) - \mathbf{E}_{F^n} S \mathbf{E}_{F^{n-1}} \log \mathbf{E}_{F_n} S \\
&\leq \mathbf{E}_{F^n} \sum_{i=1}^{n-1} \mathbf{Ent}_{F_i} S + \mathbf{E}_{F_n} (\mathbf{E}_{F^{n-1}} S \log S) - \mathbf{E}_{F^n} S \mathbf{E}_{F^{n-1}} \log \mathbf{E}_{F_n} S \\
&= \mathbf{E}_{F^n} \sum_{i=1}^{n-1} \mathbf{Ent}_{F_i} S + \mathbf{E}_{F^n} \mathbf{Ent}_{F_n} S
\end{aligned}$$

where the first inequality is obtained from the induction hypothesis, while the second and third follow by Jensen's inequality applied respectively to $\log x$ and $x \log x, x > 0$ and where we have also used Fubini theorem. □

Remark 2.2. If for each i, $\widehat{X_i}$ is an independent copy of X_i, then (by Jensen's inequality again)

$$\mathbf{Ent}_{F_i} S \leq \frac{1}{2} \mathbf{E}_{F_i \times \widehat{F_i}} (S - S_i)(\log S - \log S_i),$$

where $S_i = S(X_1, \cdots, X_{i-1}, \widehat{X_i}, X_{i+1}, \cdots, X_n)$, and $\widehat{X_i} \sim \widehat{F_i}$. Thus

$$\mathbf{Ent}_{F^n} S \leq \frac{1}{2} \mathbf{E}_{F^n} \sum_{i=1}^{n} \mathbf{E}_{F_i \times \widehat{F_i}} (S - S_i)(\log S - \log S_i). \tag{2.3}$$

Moreover since $\mathbf{E}_{F^n} \mathbf{E}_{F_i \times \widehat{F_i}} (S - S_i)(\log S - \log S_i) = \mathbf{E}_{F^n} \mathbf{E}_{\widehat{F_i}} (S - S_i)(\log S - \log S_i)$, we see that (2.3) becomes

$$\mathbf{Ent}_{F^n} S \leq \frac{1}{2} \mathbf{E}_{F^n} \sum_{i=1}^{n} \mathbf{E}_{\widehat{F_i}} (S - S_i)(\log S - \log S_i). \tag{2.4}$$

The difference between (2.3) and (2.4) is minor, it might however lead to an improvement in the method of bounded difference (or martingale difference method). It is known (and easy to verify) that if S has finite second moment, $\lim_{C \to +\infty} \mathbf{Ent}(S + C)^2 = 2\mathbf{Var} S$ and thus (2.2) is stronger (see also the next section) than the Efron–Stein inequality (in essence, the jackknife estimate of entropy is also biased upwards).

Combining the tensorization Lemma 2.1 with techniques develop in [L1], [L2] will be enough to prove our first deviation inequality. To the best of our knowledge, (2.5) below is new, it is however possible that it follows from some "martingale inequality". Recall that $S^+ = \max(S, 0)$.

Theorem 2.3. *Let $S : \mathbf{R}^n \longrightarrow \mathbf{R}$ be a statistic of the independent random variables $X_1, \cdots, X_n$ with $X_i \sim F_i$; and for each i, let $\widehat{X_i}$ be an independent copy of X_i, with*

$\widehat{X}_i \sim \widehat{F}_i$. *Then for any* $t > 0$,

$$\mathbf{P}(S - \mathbf{E}_{F^n} S \geq t) \leq \exp\left(\frac{-t^2}{4\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty}\right), \tag{2.5}$$

where $S_i = S(X_1, \cdots, X_{i-1}, \widehat{X}_i, X_{i+1}, \cdots, X_n)$, *and where* $\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty = \sup_{F^n}\left(\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right)$.

Proof. The strategy of proof is very classical. First, let us assume that in addition to the hypotheses, S is bounded and so $\mathbf{E}_{F^n} S^2 e^{\lambda S} < +\infty$, for all $\lambda > 0$. Then by Bernstein (Markov, Chebyshev, Bienaymé) inequality,

$$\mathbf{P}(S - \mathbf{E}_{F^n} S \geq t) \leq e^{-\lambda t - \lambda \mathbf{E}_{F^n} S} \mathbf{E}_{F^n} e^{\lambda S}, \tag{2.6}$$

and so we need to estimate $\mathbf{E}_{F^n} e^{\lambda S}$. To do so, and for each i, let $\widehat{X}_i$ be an independent copy of X_i with $\widehat{X}_i \sim \widehat{F}_i$. From (2.2) and (2.3), the simple inequality $(a-b)(e^a - e^b) \leq (a-b)^2(e^a + e^b)/2$, $a, b \in \mathbb{R}$, and (2.4), we get

$$\begin{aligned}
\mathbf{Ent}_{F^n} e^{\lambda S} &\leq \mathbf{E}_{F^n} \sum_{i=1}^n \mathbf{Ent}_{F_i} e^{\lambda S} \\
&\leq \frac{1}{2} \mathbf{E}_{F^n} \sum_{i=1}^n \mathbf{E}_{F_i \times \widehat{F}_i} (\lambda S - \lambda S_i)(e^{\lambda S} - e^{\lambda S_i}) \\
&\leq \frac{1}{2} \lambda^2 \mathbf{E}_{F^n} \sum_{i=1}^n \mathbf{E}_{F_i \times \widehat{F}_i} (S - S_i)^2 \frac{(e^{\lambda S} + e^{\lambda S_i})}{2} \\
&\leq \lambda^2 \mathbf{E}_{F^n} \sum_{i=1}^n \mathbf{E}_{F_i \times \widehat{F}_i} e^{\lambda S}\left((S - S_i)^+\right)^2 \\
&= \lambda^2 \mathbf{E}_{F^n} e^{\lambda S} \sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2 \\
&\leq \left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty \lambda^2 \mathbf{E}_{F^n} e^{\lambda S}
\end{aligned} \tag{2.7}$$

Setting $H(\lambda) = \lambda^{-1} \log \mathbf{E}_{F^n} e^{\lambda S}$, $\lambda \neq 0$, $H(0) = \mathbf{E}_{F^n} S$, (2.7) reads as

$$H'(\lambda) \leq \left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty.$$

Integrating this inequality gives

$$\mathbf{E}_{F^n} e^{\lambda S} \leq \exp\left(\lambda \mathbf{E}_{F^n} S + \lambda^2 \left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty\right);$$

when combined with (2.6) and a minimization in λ $\left(\lambda = \frac{t}{2\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty}\right)$ it leads to:

$$\mathbf{P}(S - \mathbf{E}_{F^n} S \geq t) \leq \exp\left(\frac{-t^2}{4\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^+\right)^2\right\|_\infty}\right),$$

for every every $t \geq 0$. To remove the extra hypothesis, truncate at $M > 0$, i.e., apply the tail result to $S \wedge M$, use the inequality $\left((S \wedge M - S_i \wedge M)^+\right)^2 \leq \left((S - S_i)^+\right)^2$, and let M tend to infinity. $\square$

Remark 2.4. (i) The presence of the positive part in (2.5) should not be surprising since only right tails estimates are provided. Essentially proceeding as above, it also follows that

$$\mathbf{P}(S - \mathbf{E}_{F^n} S \geq t) \leq \exp\left(\frac{-t^2}{2\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}(S - S_i)^2\right\|_\infty}\right). \tag{2.8}$$

However, this is a much weaker bound that the one provided by (2.5). Indeed, let $S(x) = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$ (or let S any convex function) then,

$$\left((S - S_i)^+\right)^2 \leq |X_i - \widehat{X}_i|^2 \left(\frac{\partial S}{\partial x_i}(X_1, \cdots, X_n)\right)^2;$$

and if the X_i are bounded random variables, this last inequality provides for (2.5) a bound of the right order $\exp\left(\frac{-t^2}{C\|\nabla S\|_\infty^2}\right)$. On the other hand and say for $X_1 = \cdots = X_n$, non degenerate iid random variables with $\mathbf{P}(X_1 = 0) > 0$,

$$\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}(S - S_i)^2\right\|_\infty \geq \sup_{X_1 = X_2 = \cdots = X_n = 0} \left|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}(S - S_i)^2\right| = n\mathbf{E}_{\widehat{F}_1}\widehat{X}_1^2 > 0,$$

which in (2.8) gives a weaker bound than (2.5). Also, in (2.5), there is no reason for the constant $1/4$ to be optimal.

(ii) Applying (2.5) to $-S$, gives for any $t > 0$,

$$\mathbf{P}(S - \mathbf{E}_{F^n} S \leq -t) \leq \exp\left(\frac{-t^2}{4\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F}_i}\left((S - S_i)^-\right)^2\right\|_\infty}\right). \tag{2.9}$$

Hence, two sided tails inequalities follow by adding (2.5) and (2.9) and in particular these also imply a two sided version of (2.8).

(iii) (2.5), (2.8) and (2.9) recover and complement various known results (up to universal constants). For example, if $S = X_1 + \cdots + X_n$, with $a_i \leq X_i \leq b_i$, then $(S - S_i)^2 = (X_i - \widehat{X}_i)^2$ and (2.8) implies the classical result of Hoeffding,

$$\mathbf{P}(S - \mathbf{E}S \geq t) \leq \exp\left(\frac{-t^2}{2\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{2.10}$$

Moreover, if $(S - S_i)^2 \leq c_i^2$, then the bounded difference inequality (see [GH]), also follows from (2.5). We should however mention here that since our supremum is outside the sum, we only need to have a global control on the sum but not on each individual supremum. Of course, this is irrelevant if $S = X_1 + \cdots + X_n$, or if S is invariant under any permutation of its arguments and if the X_i are iid. In this latter case, $2\left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F_i}}\left(S-S_i\right)^2\right\|_\infty = \left\|\sum_{i=1}^n \mathbf{E}_{\widehat{F_i}}\left(S_i-\overline{S}\right)^2\right\|_\infty$, which is more akin to the usual jackknife expression. Actually similar arguments allow to replace (2.5) by

$$\mathbf{P}(S-\mathbf{E}_{F^n}S \geq t) \leq \exp\left(\frac{-t^2}{4\left\|\sum_{i=1}^n \mathbf{Var}_{F_i}S\right\|_\infty}\right).$$

Indeed,

$$\begin{aligned}\mathbf{Ent}_{F_i}e^{\lambda S} &\leq \lambda\mathbf{E}_{F_i}(S-\mathbf{E}_{F_i}S)(e^{\lambda S}-e^{\lambda\mathbf{E}_{F_i}S}) \\ &\leq \lambda^2\left(\mathbf{E}_{F_i}\left((S-\mathbf{E}_{F_i}S)^+\right)^2 e^{\lambda S}+\mathbf{E}_{F_i}\left((S-\mathbf{E}_{F_i}S)^-\right)^2 e^{\lambda\mathbf{E}_{F_i}S}\right),\end{aligned}$$

hence

$$\mathbf{Ent}_{F^n}e^{\lambda S} \leq \lambda^2\mathbf{E}_{F^n}e^{\lambda S}\sum_{i=1}^n \mathbf{E}_{F_i}(S-\mathbf{E}_{F_i}S)^2,$$

and the claim follows as before. Another version of Theorem 2.3 could also have been given with $4\left\|\sum_{i=1}^n\left((S-S_i)^+\right)^2\right\|_\infty$ replacing the present denominator on the right hand side of (2.5).

(iv) It is not clear to us, how (2.5) compares to the various "martingale difference inequalities (Azuma–Hoeffding)" or their generalizations (see [GH] and the reference therein). As is well known, $S-\mathbf{E}_{F^n}S=\sum_{i=1}^n d_i$, with $d_i=\mathbf{E}(S|\mathcal{F}_i)-\mathbf{E}(S|\mathcal{F}_{i-1})$, where $\mathcal{F}_i=\sigma(X_1,\cdots,X_i)$ is the sigma–field generated by $X_1,\cdots,X_i$, $\mathcal{F}_0=\{\Omega,\emptyset\}$. Then since the d_i form a martingale difference sequence,

$$\mathbf{P}(|S-\mathbf{E}S|\geq t) \leq 2\exp\left(\frac{-t^2}{2\sum_{i=1}^n\|d_i\|_\infty^2}\right). \tag{2.11}$$

Now, we also have $d_i=\mathbf{E}(S-S_i|\mathcal{F}_i)$, and so (2.5) not only takes the positive part which can be significant but also puts the sup outside the sum rather than inside but not of $\mathbf{E}(S-S_i|\mathcal{F}_i)$ but of $(S-S_i)$. If $\mathcal{F}_i^*=\sigma(X_1,\cdots,X_{i-1},X_{i+1},\cdots,X_n)$, then the denominator of (2.8) is $2\|\sum_{i=1}^n(S-\mathbf{E}(S|\mathcal{F}_i^*))^2+\mathbf{Var}(S|\mathcal{F}_i^*)\|_\infty = 2\|\sum_{i=1}^n(S-\mathbf{E}_{F_i}S)^2+\mathbf{E}_{F_i}(S-\mathbf{E}_{F_i}S)^2\|_\infty$ (so involving conditional variance and empirical conditional variance).

(v) In (2.5), and up to a worsening of the constants, the mean can be replaced by a median or other parameters of interest, e.g., the sample mean or its bootstrap version (see, e.g., [MS] for such replacement techniques). Actually, bootstrap versions of our results also hold, e.g., (and using the notations of [E2])

$$\mathbf{P}^*(S-\mathbf{E}^*{}_{F^n}S\geq t) \leq \exp\left(\frac{-t^2}{4\left\|\sum_{i=1}^n \mathbf{E}^*{}_{\widehat{F_i}}\left((S-S_i)^+\right)^2\right\|_\infty}\right). \tag{2.12}$$

(vi) Actually, the present techniques show that

$$\mathbf{P}(S-\mathbf{E}_{F^n}S\geq t)\leq \exp\left(\frac{-t^2}{4\left\|\mathbf{E}_{\widehat{F}}\left((S-\widehat{S})^+\right)^2\right\|_\infty}\right), \tag{2.13}$$

where $\widehat{S}$ is an independent copy of S. This might be of interest when dealing with non independent X's. Finally we should also mention that, throughout, we only deal with real valued random variables but that random variables with values in an abstract space and statistics S defined on this space could have been considered.

(vii) In this note we strived for generality, it is however likely that for specific types of statistics the above results can be improved or be made more concrete. It will also be of great interest to estimates, for various classes of statistics, the quantities involved in the tail estimates.

3 Concluding Remarks

In the previous section, we obtained exponential inequalities for statistics S of the independent sample $X_1,\cdots,X_n$ using the tensorization of the entropy. In fact, many functionals tensorize ([L1]), and among those the variance. Since a main motivation for the present notes was the Efron–Stein inequality, let us recall:

Lemma 3.1 *Let $X_1,X_2,\cdots,X_n$ be independent random variables with $X_i\sim F_i$, and let the statistic $S:\mathbf{R}^n\longrightarrow\mathbf{R}$ have finite second moment. Then,*

$$\mathbf{Var}_{F^n}S\leq \mathbf{E}_{F^n}\sum_{i=1}^n\mathbf{Var}_{F_i}S, \tag{3.1}$$

where $F^n=F_1\times\cdots\times F_n$.

Proof. The proof of (3.1) is done by induction. For $n=1$, the inequality in (3.1) becomes equality and so the result is clear. Now assume the result for $n-1$, and let $F^{n-1}=F_1\times\cdots\times F_{n-1}$. Then we have

$$\begin{aligned}
\mathbf{Var}_{F^n}S &= \mathbf{E}_{F^n}S^2-(\mathbf{E}_{F^n}S)^2\\
&= \mathbf{E}_{F_n}\mathbf{Var}_{F^{n-1}}S+\mathbf{E}_{F_n}(\mathbf{E}_{F^{n-1}}S)^2-(\mathbf{E}_{F^n}S)^2\\
&\leq \mathbf{E}_{F_n}\mathbf{E}_{F^{n-1}}\sum_{i=1}^{n-1}\mathbf{Var}_{F_i}S+\mathbf{E}_{F_n}(\mathbf{E}_{F^{n-1}}S-\mathbf{E}_{F^{n-1}}\mathbf{E}_{F_n}S)^2\\
&= \mathbf{E}_{F^n}\sum_{i=1}^{n-1}\mathbf{Var}_{F_i}S+\mathbf{E}_{F_n}(\mathbf{E}_{F^{n-1}}(S-\mathbf{E}_{F_n}S))^2\\
&\leq \mathbf{E}_{F^n}\sum_{i=1}^{n-1}\mathbf{Var}_{F_i}S+\mathbf{E}_{F_n}(\mathbf{E}_{F^{n-1}}(S-\mathbf{E}_{F_n}S)^2)\\
&= \mathbf{E}_{F^n}\sum_{i=1}^{n-1}\mathbf{Var}_{F_i}S+\mathbf{E}_{F^n}\mathbf{Var}_{F_n}S
\end{aligned}$$

where the first inequality is obtained from the induction hypothesis and the second by Cauchy–Schwarz, and where we have also used Fubini theorem. $\square$

Remark 3.2. (i) If for each i, $\widehat{X_i}$ is an independent copy of X_i, then (3.1) becomes

$$\mathbf{Var}_{F^n} S \leq \frac{1}{2} \mathbf{E}_{F^n} \sum_{i=1}^{n} \mathbf{E}_{F_i \times \widehat{F_i}} (S - S_i)^2, \tag{3.2}$$

where $S_i = S(X_1, \cdots, X_{i-1}, \widehat{X_i}, X_{i+1}, \cdots, X_n)$. Again, since $\mathbf{E}_{F^n} \mathbf{E}_{F_i \times \widehat{F_i}} (S - S_i)^2 = \mathbf{E}_{F^n} \mathbf{E}_{\widehat{F_i}} (S - S_i)^2$, (3.1) can also be written as

$$\mathbf{Var}_{F^n} S \leq \frac{1}{2} \mathbf{E}_{F^n} \sum_{i=1}^{n} \mathbf{E}_{\widehat{F_i}} (S - S_i)^2. \tag{3.3}$$

(ii) If S is symmetric, i.e., invariant under any permutation of its arguments, and if $X_1, \cdots, X_n, \widehat{X}_{n+1}$ are iid, (3.1)–(3.3) recover the Efron-Stein inequality (see [ES]). Still for iid samples but without any symmetry assumption, (3.1) also recovers an inequality of Steele ([S]). Moreover, (3.1)–(3.3) is more in line with the jackknife methodology which requires resampling a single copy of a potential outlier. In contrast to the published proofs of these inequalities, the proof of (3.1) does not require an orthogonal decomposition method (see [RT], [V1], [V2] and [H] for further details and more references). Finally, it is clear that equality holds in (3.1)–(3.3) if and only if the the statistic S is linear, i.e., $S - \mathbf{E}_{F^n} S = \sum_{i=1}^{n} f_i(X_i)$.

Acknowledgments: Many thanks to R. Latała for his comments on an early version of this paper.

References

[E1] Efron, B. (1979), Bootstrap Methods: Another look at the Jackknife. *Ann. Statist.* **7**, 1–26.

[E2] Efron, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans SIAM, Philadelphia.

[ET] Efron, B. and Tibsirani, R. (1993), An introduction to the Bootstrap. Chapman and Hall, New York.

[ES] Efron, B. and Stein, C. (1981), The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

[F] Frangos, C.C. (1987), An updated bibliography on the jackknife method. *Comm. Statist. Theory Metods* **16**, 1543–1584.

[GH] Godbole, A.P. and Hitczenko, P. (1998), Beyond the method of bounded differences. Proceedings of the DIMACS Workshop "Microsurveys in Discrete Probability" DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **41**, 43–58.

[H] Houdré, C. (1997), The iterated Jackknife estimate of variance. *Stat. and Probab. Letters* **35**, 197–201.

[HT] Houdré, C. and Tetali, P. (1997), Concentration of measure for products of Markov kernels via functional inequalities, preprint.

[L1] Ledoux, M. (1996), On Talagrand's deviation inequalities for product measures, *ESAIM Prob. & Stat.*, **1**, 63-87.

[L2] Ledoux, M. (1997), Concentration of measure and Logarithmic Sobolev inequalities. Preprint.

[L] Lieb, E. (1975), Some convexity and subadditivity properties of entropy. *Bull. Amer. Math. Soc.* **81**, 1–13.

[M] Miller, R.G. (1974), The Jackknife–A Review. *Biometrika* **61**, 1–15.

[MS] Milman, V. D. and Schechtman, G.. Asymptotic theory of finite dimensional normed spaces. Lecture Notes in Math. **1200** (1986), Springer-Verlag.

[PS] Parr, W.C. and Schucany, W.R. (1980), The Jackknife: A Bibliography. *Internat. Statist. rev.* **48**, 73–78.

[RT] Rhee, W.T. and Talagrand, M. (1986), Martingale Inequalities and the jackknife estimate of variance. *Statist. Probab. Letters* **4** 5–6.

[S] Steele, J.M. (1986), An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.* **14**, 753–758.

[V1] Vitale, R.A. (1984), An expansion for symmetric statistics and the Efron-Stein inequality. In *Inequalities in Statistics and Probability*, IMS Lecture Notes–Monograph Series **5** (Y.L. Tong, ed.) 112–114, IMS, Hayward, Calif.

[V2] Vitale, R.A. (1993), On the bias of jackknife estimate of variance. In *Stochastic Inequalities*, IMS Lecture Notes–Monograph Series **22** (M. Shaked and Y.L. Tong, eds.) 399–403, IMS, Hayward, Calif.