Hindawi Publishing Corporation
Journal of Inequalities and Applications
Volume 2009, Article ID 941936, 20 pages
doi:10.1155/2009/941936

Research Article

Bounds for Tail Probabilities of the Sample Variance

V. Bentkus[1] and M. Van Zuijlen[2]

[1] *Vilnius Pedagogical University, Studentu 39, LT-08106 Vilnius, Lithuania*
[2] *IMAPP, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*

Correspondence should be addressed to V. Bentkus, bentkus@ktl.mii.lt

Received 11 February 2009; Accepted 20 June 2009

Recommended by Andrei Volodin

We provide bounds for tail probabilities of the sample variance. The bounds are expressed in terms of Hoeffding functions and are the sharpest known. They are designed having in mind applications in auditing as well as in processing data related to environment.

1. Introduction and Results

Let $X, X_1, \dots, X_n$ be a random sample of independent identically distributed observations. Throughout we write

$$\mu = \mathbb{E}X, \quad \sigma^2 = \mathbb{E}(X-\mu)^2, \quad \omega = \mathbb{E}(X-\mu)^4 \tag{1.1}$$

for the mean, variance, and the fourth central moment of X, and assume that $n \geq 2$. Some of our results hold only for bounded random variables. In such cases without loss of generality we assume that $0 \leq X \leq 1$. Note that $0 \leq X \leq 1$ is a natural condition in audit applications.

The sample variance $\widehat{\sigma}^2$ of the sample $X_1, \dots, X_n$ is defined as

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \tag{1.2}$$

where $\overline{X}$ is the sample mean, $n\overline{X} = X_1 + \cdots + X_n$. We can rewrite (1.2) as

$$\widehat{\sigma}^2 = \frac{1}{n(n-1)} \sum_{i \neq j,\ 1 \leq i,j \leq n} \frac{(X_i - X_j)^2}{2}. \tag{1.3}$$

We are interested in deviations of the statistic $\widehat{\sigma}^2$ from its mean $\sigma^2 = \mathbb{E}\widehat{\sigma}^2$, that is, in bounds for the tail probabilities of the statistic $T = \sigma^2 - \widehat{\sigma}^2$,

$$\mathbb{P}\{T \geq t\} = \mathbb{P}\left\{\widehat{\sigma}^2 \leq \sigma^2 - t\right\}, \quad 0 \leq t \leq \sigma^2, \tag{1.4}$$

$$\mathbb{P}\{T \leq -t\} = \mathbb{P}\left\{\widehat{\sigma}^2 \geq \sigma^2 + t\right\}, \quad t \geq 0. \tag{1.5}$$

The paper is organized as follows. In the introduction we give a description of bounds, some comments, and references. In Section 2 we obtain sharp upper bounds for the fourth moment. In Section 3 we give proofs of all facts and results from the introduction.

If $0 \leq X \leq 1$, then the range of interest in (1.5) is $0 \leq t \leq \gamma^2$, where

$$\gamma^2 = \begin{cases} \dfrac{1}{4} - \sigma^2 + \dfrac{1}{4(n-1)}, & \text{if } n \text{ is even}, \\[2ex] \dfrac{1}{4} - \sigma^2 + \dfrac{1}{4n}, & \text{if } n \text{ is odd}. \end{cases} \tag{1.6}$$

The restriction $0 \leq t \leq \sigma^2$ on the range of t in (1.4) (resp., $0 \leq t \leq \gamma^2$ in (1.5) in cases where the condition $0 \leq X \leq 1$ is fulfilled) is natural. Indeed, $P\{T \geq t\} = 0$ for $t > \sigma^2$, due to the obvious inequality $\widehat{\sigma}^2 \geq 0$. Furthermore, in the case of $0 \leq X \leq 1$ we have $\mathbb{P}\{T \leq -t\} = 0$ for $t > \gamma^2$ since $\widehat{\sigma}^2 \leq \gamma^2 + \sigma^2$ (see Proposition 2.3 for a proof of the latter inequality).

The asymptotic (as $n \to \infty$) properties of T (see Section 3 for proofs of (1.7) and (1.8)) can be used to test the quality of bounds for tail probabilities. Under the condition $\mathbb{E}X^4 < \infty$ the statistic $T = \sigma^2 - \widehat{\sigma}^2$ is asymptotically normal provided that X is not a Bernoulli random variable symmetric around its mean. Namely, if $\omega > \sigma^4$, then

$$\lim_{n \to \infty} \mathbb{P}\left\{\sqrt{n}T \geq y\sqrt{\omega - \sigma^4}\right\} = 1 - \Phi(y), \quad y \in \mathbb{R}. \tag{1.7}$$

If $\omega = \sigma^4$ (which happens if and only if X is a Bernoulli random variable symmetric around its mean), then asymptotically T has χ^2 type distribution, that is,

$$\lim_{n \to \infty} \mathbb{P}\left\{nT \geq y\sigma^2\right\} = \mathbb{P}\left\{\eta^2 - 1 \geq y\right\}, \quad y \in \mathbb{R}, \tag{1.8}$$

where η is a standard normal random variable, and $\Phi(y) = \mathbb{P}\{\eta \leq y\}$ is the standard normal distribution function.

Let us recall already known bounds for the tail probabilities of the sample variance (see (1.19)–(1.21)). We need notation related to certain functions coming back to Hoeffding [1]. Let $0 < p \le 1$ and $q = 1 - p$. Write

$$H(x;p) = \left(1 + \frac{qx}{p}\right)^{-qx-p} (1-x)^{qx-q}, \quad 0 \le x \le 1. \tag{1.9}$$

For $x \le 0$ we define $H(x;p) = 1$. For $x > 1$ we set $H(x;p) = 0$. Note that our notation for the function H is slightly different from the traditional one. Let $\lambda \ge 0$. Introduce as well the function

$$\Pi(x;\lambda) = e^x \left(1 + \frac{x}{\lambda}\right)^{-x-\lambda} \quad \text{for } x \ge 0, \tag{1.10}$$

and $\Pi(x;\lambda) = 1$ for $x \le 0$. One can check that

$$H(x;p) \le \Pi\left(x; \frac{p}{q}\right). \tag{1.11}$$

All our bounds are expressed in terms of the function H. Using (1.11), it is easy to replace them by bounds expressed in terms of the function Π, and we omit related formulations.

Let $0 \le p < 1$ and $\sigma^2 \ge 0$. Assume that

$$p = \frac{\sigma^2}{1+\sigma^2}, \quad q = \frac{1}{1+\sigma^2}, \quad p + q = 1. \tag{1.12}$$

Let ε be a Bernoulli random variable such that $\mathbb{P}\{\varepsilon = -\sigma^2\} = q$ and $\mathbb{P}\{\varepsilon = 1\} = p$. Then $\mathbb{E}\varepsilon = 0$ and $\mathbb{E}\varepsilon^2 = \sigma^2$. The function H is related to the generating function (the Laplace transform) of binomial distributions since

$$H(x;p) = \inf_{h>0} \exp\{-hx\} \mathbb{E} \exp\{h\varepsilon\}, \tag{1.13}$$

$$H^n(x;p) = \inf_{h>0} \exp\{-hnx\} \mathbb{E} \exp\{h(\varepsilon_1 + \cdots + \varepsilon_n)\}, \tag{1.14}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent copies of ε. Note that (1.14) is an obvious corollary of (1.13). We omit elementary calculations leading to (1.13). In a similar way

$$\Pi(x;\lambda) = \inf_{h>0} \exp\{-hx\} \mathbb{E} \exp\{h(\eta - \lambda)\}, \tag{1.15}$$

where η is a Poisson random variable with parameter λ.

The functions H and Π satisfy a kind of the Central Limit Theorem. Namely, for given $0 < p < 1$ and $y \ge 0$ we have

$$\lim_{n\to\infty} H^n\left(yn^{-1/2}\sqrt{\frac{p}{q}}; p\right) = \lim_{n\to\infty} \Pi^n\left(yn^{-1/2}\sqrt{\lambda}; \lambda\right) = \exp\left\{-\frac{y^2}{2}\right\} \tag{1.16}$$

(we omit elementary calculations leading to (1.16)). Furthermore, we have [1]

$$H\left(y\sqrt{\frac{p}{q}};p\right) \le \exp\left\{-\frac{y^2}{2}\right\}, \quad \frac{1}{2} \le p < 1,\ y \ge 0, \tag{1.17}$$

and we also have [2]

$$H\left(\frac{yp}{q};p\right) \le \exp\left\{-\frac{py^2}{2q(y+1)}\right\}, \quad 0 \le p \le \frac{1}{2},\ y \ge 0. \tag{1.18}$$

Using the introduced notation, we can recall the known results (see [2, Lemma 3.2]). Let $k = [n/2]$ be the integer part of $n/2$. Assume that $0 \le X \le 1$. If σ^2 is known, then

$$\mathbb{P}\{T \ge t\} \le U_0, \quad U_0 \overset{\text{def}}{=} H^k\left(\frac{t}{\sigma^2};1-2\sigma^2\right). \tag{1.19}$$

The right-hand side of (1.19) is an increasing function of $\sigma^2 \le 1/4$ (see Section 3 for a short proof of (1.19) as a corollary of Theorem 1.1). If σ^2 is unknown but μ is known, then

$$\mathbb{P}\{T \ge t\} \le U_1, \quad U_1 \overset{\text{def}}{=} H^k\left(\frac{t}{\mu-\mu^2};1-2\mu+2\mu^2\right). \tag{1.20}$$

Using the obvious estimate $\sigma^2 \le \mu(1-\mu)$, the bound (1.20) is implied by (1.19). In cases where both μ and σ^2 are not known, we have

$$\mathbb{P}\{T \ge t\} \le U_2, \quad U_2 \overset{\text{def}}{=} H^k\left(4t;\frac{1}{2}\right), \tag{1.21}$$

as it follows from (1.19) using the obvious bound $\sigma^2 \le 1/4$.

Let us note that the known bounds (1.19)–(1.21) are the best possible in the framework of an approach based on analysis of the variance, usage of exponential functions, and of an inequality of Hoeffding (see (3.3)), which allows to reduce the problem to estimation of tail probabilities for sums of independent random variables. Our improvement is due to careful analysis of the fourth moment which appears to be quite complicated; see Section 2. Briefly the results of this paper are the following: we prove a general bound involving μ, σ^2, and the fourth moment ω; this general bound implies all other bounds, in particular a new precise bound involving μ and σ^2; we provide as well bounds for lower tails $\mathbb{P}\{T \le -t\}$; we compare the bounds analytically, mostly as n is sufficiently large.

From the mathematical point of view the sample variance is one of the simplest nonlinear statistics. Known bounds for tail probabilities are designed having in mind linear statistics, possibly also for dependent observations. See a seminal paper of Hoeffding [1] published in JASA. For further development see Talagrand [3], Pinelis [4, 5], Bentkus [6, 7], Bentkus et al. [8, 9], and so forth. Our intention is to develop tools useful in the setting of nonlinear statistics, using the sample variance as a test statistic.

Theorem 1.1 extends and improves the known bounds (1.19)–(1.21). We can derive (1.19)–(1.21) from this theorem since we can estimate the fourth moment ω via various combinations of μ and σ^2 using the boundedness assumption $0 \le X \le 1$.

Theorem 1.1. *Let $k = [n/2]$ and $\omega_0 \ge 0$.*

If $\mathbb{E}X^4 < \infty$ and $\omega \le \omega_0$, then

$$\mathbb{P}\{T \ge t\} \le U, \quad U \overset{\text{def}}{=} H^k\left(\frac{t}{\sigma^2}; p\right) \tag{1.22}$$

with

$$p = \frac{\sigma^4 + \omega_0}{3\sigma^4 + \omega_0} = \frac{s^2}{1 + s^2}, \quad s^2 = \frac{\sigma^4 + \omega_0}{2\sigma^4}. \tag{1.23}$$

If $0 \le X \le 1$ and $\omega \le \omega_0$, then

$$\mathbb{P}\{T \le -t\} \le L, \quad L \overset{\text{def}}{=} H^k\left(\frac{2t}{1 - 2\sigma^2}; p\right) \tag{1.24}$$

with

$$p = \frac{2\sigma^4 + 2\omega_0}{1 - 4\sigma^2 + 6\sigma^4 + 2\omega_0} = \frac{s^2}{1 + s^2}, \quad s^2 = \frac{2\sigma^4 + 2\omega_0}{(1 - 2\sigma^2)^2}. \tag{1.25}$$

Both bounds U and L are increasing functions of p, ω_0, and s^2.

Remark 1.2. In order to derive upper confidence bounds we need only estimates of the upper tail $\mathbb{P}\{T \ge t\}$ (see [2]). To estimate the upper tail the condition $\mathbb{E}X^4 < \infty$ is sufficient. The lower tail $\mathbb{P}\{T \le -t\}$ has a different type of behavior since to estimate it we indeed need the assumption that X is a bounded random variable.

For $0 \le X \le 1$ Theorem 1.1 implies the known bounds (1.19)–(1.21) for the upper tail of T. It implies as well the bounds (1.26)–(1.29) for the lower tail. The lower tail has a bit more complicated structure, (cf. (1.26)–(1.29) with their counterparts (1.19)–(1.21) for the upper tail).

If σ^2 is known, then

$$\mathbb{P}\{T \le -t\} \le L_0, \quad L_0 \overset{\text{def}}{=} H^k\left(\frac{2t}{1 - 2\sigma^2}; 2\sigma^2\right). \tag{1.26}$$

One can show (we omit details) that the bound L_0 is not an increasing function of σ^2. A bit rougher inequality

$$\mathbb{P}\{T \le -t\} \le L_0^*, \quad L_0^* \overset{\text{def}}{=} H^k\left(2t; \frac{2\sigma^2}{1 + 2\sigma^2}\right) \tag{1.27}$$

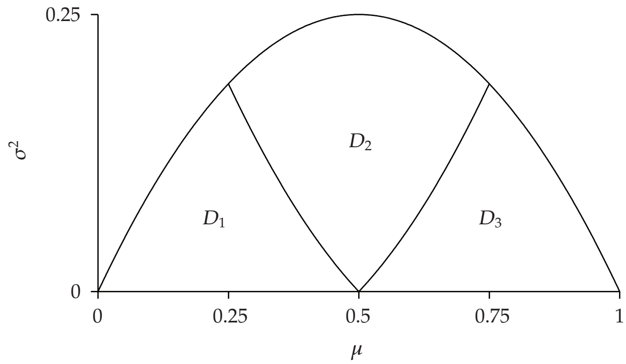

Figure 1: $D = D_1 \cup D_2 \cup D_3$.

has the monotonicity property since L_0^* is an increasing function of σ^2. If μ is known, then using the obvious inequality $\sigma^2 \le \mu(1-\mu)$, the bound (1.27) yields

$$\mathbb{P}\{T \le -t\} \le L_1, \quad L_1 \stackrel{\text{def}}{=} H^k\left(2t; \frac{2\mu - 2\mu^2}{1 + 2\mu - 2\mu^2}\right). \tag{1.28}$$

If we have no information about μ and σ^2, then using $\sigma^2 \le 1/4$, the bound (1.27) implies

$$\mathbb{P}\{T \le -t\} \le L_2, \quad L_2 \stackrel{\text{def}}{=} H^k\left(2t; \frac{1}{3}\right). \tag{1.29}$$

The bounds above do not cover the situation where both μ and σ^2 are known. To formulate a related result we need additional notation. In case of $0 \le X \le 1$ we use the notation

$$f_1 = (1-\mu)\left(\frac{1}{2} - \mu\right), \qquad f_3 = \mu\left(\mu - \frac{1}{2}\right). \tag{1.30}$$

In view of the well-known upper bound $\sigma^2 \le \mu(1-\mu)$ for the variance of $0 \le X \le 1$, we can partition the set

$$D = \left\{\left(\mu, \sigma^2\right) \in \mathbb{R}^2 : 0 \le \mu \le 1, 0 \le \sigma^2 \le \mu(1-\mu)\right\} \tag{1.31}$$

of possible values of μ and σ^2 into a union $D = D_1 \cup D_2 \cup D_3$ of three subsets

$$D_1 = \left\{\left(\mu, \sigma^2\right) \in D : \sigma^2 \le f_1\right\}, \qquad D_3 = \left\{\left(\mu, \sigma^2\right) \in D : \sigma^2 \le f_3\right\}, \tag{1.32}$$

and $D_2 = D \setminus (D_1 \cup D_3)$; see Figure 1.

Theorem 1.3. *Write* $k = [n/2]$. *Assume that* $0 \le X \le 1$.

The upper tail of the statistic T satisfies

$$\mathbb{P}\{T \geq t\} \leq U_3, \quad U_3 \overset{\text{def}}{=} H^k\left(\frac{t}{\sigma^2}; p_u\right) \tag{1.33}$$

with $p_u = s^2/(1+s^2)$, *where*

$$s^2 = \begin{cases} \dfrac{\sigma^4 + (1-\mu)^4}{2(1-\mu)^2\sigma^2}, & \text{if } (\mu, \sigma^2) \in D_1, \\ \dfrac{a + b\sigma^2 + 4\sigma^4}{8\sigma^4}, & \text{if } (\mu, \sigma^2) \in D_2, \\ \dfrac{\sigma^4 + \mu^4}{2\mu^2\sigma^2}, & \text{if } (\mu, \sigma^2) \in D_3, \end{cases} \tag{1.34}$$

and where one can write

$$a = \mu(1-\mu)(2\mu-1)^2, \qquad b = 8\mu^2 - 8\mu + 3. \tag{1.35}$$

The lower tail of T satisfies

$$\mathbb{P}\{T \leq -t\} \leq L_3, \quad L_3 \overset{\text{def}}{=} H^k\left(\frac{2t}{1-2\sigma^2}; p_l\right) \tag{1.36}$$

with $p_l = s^2/(c^2+s^2)$, *where* $c = (1-2\sigma^2)/(2\sigma^2)$, *and* s^2 *is defined by* (1.34).

The bounds above are obtained using the classical transform $G \mapsto HG$,

$$(HG)(x) = \inf_{h<x} \mathbb{E} \exp\{h(Y-x)\} \tag{1.37}$$

of survival functions $G(x) = \mathbb{P}\{Y \geq x\}$ (cf. definitions (1.13) and (1.14) of the related Hoeffding functions). The bounds expressed in terms of Hoeffding functions have a simple analytical structure and are easily numerically computable.

All our upper and lower bounds satisfy a kind of the Central Limit Theorem. Namely, if we consider an upper bound, say $U = U(t)$ (resp., a lower bound $L = L(t)$) as a function of t, then there exist limits

$$\lim_{n\to\infty} U\left(\frac{t}{\sqrt{n}}\right) = \exp\left\{-ct^2\right\}, \qquad \lim_{n\to\infty} L\left(\frac{t}{\sqrt{n}}\right) = \exp\left\{-dt^2\right\} \tag{1.38}$$

with some positive c and d. The values of c and d can be used to compare the bounds—the larger these constants, the better the bound. To prove (1.38) it suffices to note that with $k = [n/2]$

$$\lim_{n\to\infty} H^k\left(\frac{x}{\sqrt{n}}; p\right) = \exp\left\{-\frac{qx^2}{4p}\right\}. \tag{1.39}$$

The Central Limit Theorem in the form of (1.7) restricts the ranges of possible values of c and d. Namely, using (1.7) it is easy to see that c and d have to satisfy

$$c, d \leq a \stackrel{\text{def}}{=} \frac{1}{2(\omega - \sigma^4)}. \tag{1.40}$$

We provide the values of these constants for all our bounds and give the numerical values of them in the following two cases.

(i) X is a random variable uniformly distributed in the interval $[0, 1/2]$. The moments of this random variable satisfy

$$\mu = \frac{1}{4}, \quad \sigma^2 = \frac{1}{48}, \quad (\mu, \sigma^2) \in D_1, \qquad \omega = \frac{1}{1280}, \qquad a = 1440. \tag{1.41}$$

For μ, σ^2, ω defined by (1.41), the constants c and d we give as c_1, d_1.

(ii) X is uniformly distributed in $[0, 1]$, and in this case

$$\mu = \frac{1}{2}, \quad \sigma^2 = \frac{1}{12}, \quad (\mu, \sigma^2) \in D_2, \qquad \omega = \frac{1}{80}, \qquad a = 90. \tag{1.42}$$

For the constants c and d with μ, σ^2, ω defined by (1.42) we give as c_2, d_2.

We have

$$\begin{aligned}
&U_2\colon \quad c = 4, \qquad c_1 = 4, \qquad c_2 = 4, \\
&U_1\colon \quad c = \frac{1}{(2\mu - 2\mu^2)(1 - 2\mu + 2\mu^2)}, \qquad c_1 = 4.26\ldots, \qquad c_2 = 4, \\
&U_0\colon \quad c = \frac{1}{2\sigma^2 - 4\sigma^4}, \qquad c_1 = 25.04\ldots, \qquad c_2 = 7.2, \\
&U_3\colon \quad c = \frac{1}{4\sigma^4 s^2}, \qquad c_1 = 42.60\ldots, \qquad c_2 = 18,
\end{aligned} \tag{1.43}$$

$$U: \quad c = \frac{1}{2\sigma^4 + 2\omega_0}, \qquad c_1 = 411.42\ldots, \qquad c_2 = 25.71\ldots, \tag{1.44}$$

$$L_2: \quad d = 2, \qquad d_1 = 2, \qquad d_2 = 2,$$

$$L_1: \quad d = \frac{1}{2\mu - 2\mu^2}, \qquad d_1 = 2.66\ldots, \qquad d_2 = 2,$$

$$L_0^*: \quad d = \frac{1}{2\sigma^2}, \qquad d_1 = 24, \qquad d_2 = 6, \tag{1.45}$$

$$L_0: \quad d = \frac{1}{2\sigma^2 - 4\sigma^4}, \qquad d_1 = 25.04\ldots, \qquad d_2 = 7.2,$$

$$L_3: \quad d = \frac{1}{4\sigma^4 s^2}, \qquad d_1 = 42.60\ldots, \qquad d_2 = 18,$$

$$L: \quad d = \frac{1}{2\sigma^4 + 2\omega_0}, \qquad d_1 = 411.42\ldots, \qquad d_2 = 25.71\ldots, \tag{1.46}$$

while calculating the constants in (1.44) and (1.46) we choose $\omega_0 = \omega$. The quantity s^2 in (1.43) and (1.45) is defined by (1.34).

Conclusions

Our new bounds provide a substantial improvement of the known bounds. However, from the asymptotic point of view these bounds seem to be still rather crude. To improve the bounds further one needs new methods and approaches. Some preliminary computer simulations show that in applications where n is finite and random variables have small means and variances (like in auditing, where a typical value of n is 60), the asymptotic behavior is not related much to the behavior for small n. Therefore bounds specially designed to cover the case of finite n have to be developed.

2. Sharp Upper Bounds for the Fourth Moment

Recall that we consider bounded random variables such that $0 \le X \le 1$, and that we write $\mu = \mathbb{E}X$ and $\sigma^2 = \mathbb{E}(X-\mu)^2$. In Lemma 2.1 we provide an optimal upper bound for the fourth moment of $X - \lambda$ given a shift $\lambda \in \mathbb{R}$, a mean μ, and a variance σ^2. The maximizers of the fourth moment are either Bernoulli or trinomial random variables. It turns out that their distributions, say ν, are of the following three types (i)–(iii):

(i) a two point distribution such that

$$\nu(\{d\}) = r, \qquad \nu(\{1\}) = p, \qquad d = \mu - \frac{\sigma^2}{1-\mu}, \tag{2.1}$$

$$r = \frac{(1-\mu)^2}{(1-\mu)^2 + \sigma^2}, \qquad p = \frac{\sigma^2}{(1-\mu)^2 + \sigma^2}; \tag{2.2}$$

(ii) a family of three point distributions depending on $1/4 < \lambda < 3/4$ such that

$$\nu(\{0\}) = q, \qquad \nu(\{d\}) = r, \qquad \nu(\{1\}) = p, \qquad d \equiv d_\lambda = 2\lambda - \frac{1}{2}, \tag{2.3}$$

$$q = \frac{\sigma^2 - f_1}{d_\lambda}, \qquad r = \frac{\mu(1-\mu) - \sigma^2}{d_\lambda(1-d_\lambda)}, \qquad p = \frac{\sigma^2 - f_3}{1 - d_\lambda}, \tag{2.4}$$

where we write

$$f_1 = (1-\mu)(\mu - d_\lambda), \qquad f_3 = \mu(d_\lambda - \mu); \tag{2.5}$$

notice that (2.4) supplies a three-point probability distribution only in cases where the inequalities $\sigma^2 > f_1$ and $\sigma^2 > f_3$ hold;

(iii) a two point distribution such that

$$\nu(\{0\}) = q, \qquad \nu(\{d\}) = r, \qquad d = \mu + \frac{\sigma^2}{\mu}, \tag{2.6}$$

$$q = \frac{\sigma^2}{\mu^2 + \sigma^2}, \qquad r = \frac{\mu^2}{\mu^2 + \sigma^2}. \tag{2.7}$$

Note that the point d in (2.2)–(2.7) satisfies $0 \le d \le 1$ and that the probability distribution ν has mean μ and variance σ^2.

Introduce the set

$$D = \left\{ \left(\mu, \sigma^2\right) \in \mathbb{R}^2 : \mu = \mathbb{E}X,\ \sigma^2 = \mathbb{E}(X-\mu)^2,\ 0 \le X \le 1 \right\}. \tag{2.8}$$

Using the well-known bound $\sigma^2 \le \mu(1-\mu)$ valid for $0 \le X \le 1$, it is easy to see that

$$D = \left\{ \left(\mu, \sigma^2\right) \in \mathbb{R}^2 : 0 \le \mu \le 1,\ 0 \le \sigma^2 \le \mu(1-\mu) \right\}. \tag{2.9}$$

Let $\lambda \in \mathbb{R}$. We represent the set $D \subset \mathbb{R}^2$ as a union $D = D_1^\lambda \cup D_2^\lambda \cup D_3^\lambda$ of three subsets setting

$$D_1^\lambda = \left\{ \left(\mu, \sigma^2\right) \in D : \sigma^2 \le f_1 \right\}, \qquad D_3^\lambda = \left\{ \left(\mu, \sigma^2\right) \in D : \sigma^2 \le f_3 \right\}, \tag{2.10}$$

and $D_2^\lambda = D \setminus (D_1^\lambda \cup D_3^\lambda)$, where f_1 and f_3 are given in (2.5). Let us mention the following properties of the regions.

(a) If $\lambda \le 1/4$, then $D = D_1^\lambda$ since for such λ obviously $\mu(1-\mu) \le f_1$ for all $0 \le \mu \le 1$. The set $D_3^\lambda = \{(0,0)\}$ is a one-point set. The set D_2^λ is empty.

(b) If $\lambda \ge 3/4$, then $D = D_3^\lambda$ since for such λ clearly $\mu(1-\mu) \le f_3$ for all $0 \le \mu \le 1$. The set $D_1^\lambda = \{(1,0)\}$ is a one-point set. The set D_2^λ is empty.

For $1/4 < \lambda < 3/4$ all three regions D_1^λ, D_2^λ, D_3^λ are nonempty sets. The sets D_1^λ and D_3^λ have only one common point $(d_\lambda, 0) \in D$, that is, $D_1^\lambda \cap D_3^\lambda = \{(d_\lambda, 0)\}$.

Lemma 2.1. *Let $\lambda \in \mathbb{R}$. Assume that a random variable X satisfies*

$$0 \le X \le 1, \quad \mathbb{E}X = \mu, \quad \mathbb{E}(X-\mu)^2 = \sigma^2. \tag{2.11}$$

Then

$$\mathbb{E}(X-\lambda)^4 \le \mathbb{E}(X_*-\lambda)^4 \tag{2.12}$$

with a random variable X_* satisfying (2.11) *and defined as follows:*

(i) *if $(\mu, \sigma^2) \in D_1^\lambda$, then X_* is a Bernoulli random variable with distribution* (2.2)*;*

(ii) *if $(\mu, \sigma^2) \in D_2^\lambda$, then X_* is a trinomial random variable with distribution* (2.4)*;*

(iii) *if $(\mu, \sigma^2) \in D_3^\lambda$, then X_* is a Bernoulli random variable with distribution* (2.7).

Proof. Writing $Y = X - \lambda$, we have to prove that if

$$-\lambda \le Y \le 1-\lambda, \quad \mathbb{E}Y = \mu - \lambda, \quad \mathbb{E}(Y - \mathbb{E}Y)^2 = \sigma^2, \tag{2.13}$$

then

$$\mathbb{E}Y^4 \le \mathbb{E}Y_*^4 \tag{2.14}$$

with $Y_* = X_* - \lambda$. Henceforth we write $a = d - \lambda$, so that Y_* can assume only the values $-\lambda$, a, $1-\lambda$ with probabilities q, r, p defined in (2.2)–(2.7), respectively. The distribution $\varrho = \mathcal{L}(Y_*)$ is related to the distribution $\nu = \mathcal{L}(X_*)$ as $\varrho(B) = \nu(B + \lambda)$ for all $B \subset \mathbb{R}$.

Formally in our proof we do not need the description (2.17) of measures ϱ satisfying (2.15). However, the description helps to understand the idea of the proof. Let $a \in \mathbb{R}$ and $\sigma^2 \ge 0$. Assume that a signed measure ϱ of subsets of $\mathbb{R}$ is such that the total variation measure $\varrho_+ + \varrho_-$ is a discrete measure concentrated in a three-point set $\{-\lambda, a, 1-\lambda\}$ and

$$\int_{\mathbb{R}} \varrho(dx) = 1, \qquad \int_{\mathbb{R}} x\varrho(dx) = \mu - \lambda, \qquad \int_{\mathbb{R}} (x - \mu + \lambda)^2 \varrho(dx) = \sigma^2. \tag{2.15}$$

Then ϱ is a uniquely defined measure such that

$$q \overset{\text{def}}{=} \varrho(\{-\lambda\}), \qquad r \overset{\text{def}}{=} \varrho(\{a\}), \qquad p \overset{\text{def}}{=} \varrho(\{1-\lambda\}) \tag{2.16}$$

satisfy

$$q = \frac{\sigma^2 + (a - \mu + \lambda)(1-\mu)}{a + \lambda}, \qquad r = \frac{\mu(1-\mu) - \sigma^2}{(a+\lambda)(1-a-\lambda)}, \qquad p = \frac{\sigma^2 - (a - \mu + \lambda)\mu}{1 - a - \lambda}. \tag{2.17}$$

We omit the elementary calculations leading to (2.17). The calculations are related to solving systems of linear equations.

Let $a, b, c \in \mathbb{R}$. Consider the polynomial

$$P(t) = (t-c)(b-t)(t-a)^2 \equiv c_0 + c_1 t + c_2 t^2 + c_3 t^3 - t^4, \quad t \in \mathbb{R}. \tag{2.18}$$

It is easy to check that

$$c_3 = 0 \Longleftrightarrow b + c + 2a = 0. \tag{2.19}$$

The proofs of (i)–(iii) differ only in technical details. In all cases we find a, b, and c (depending on λ, μ and σ^2) such that the polynomial P defined by (2.18) satisfies $P(t) \geq 0$ for $-\lambda \leq t \leq 1-\lambda$, and such that the coefficient c_3 in (2.18) vanishes, $c_3 = 0$. Using $c_3 = 0$, the inequality $P(t) \geq 0$ is equivalent to $t^4 \leq c_0 + c_1 t + c_2 t^2$, which obviously leads to $\mathbb{E}Y^4 \leq c_0 + c_1(\mu - \lambda) + c_2\sigma^2$. We note that the random variable Y_* assumes the values from the set

$$\{t : P(t) = 0\} = \left\{t : c_0 + c_1 t + c_2 t^2 = t^4\right\}. \tag{2.20}$$

Therefore we have

$$\mathbb{E}Y^4 \leq c_0 + c_1(\mu - \lambda) + c_2\sigma^2 = \mathbb{E}Y_*^4, \tag{2.21}$$

which proves the lemma.

(i) Now $(\mu, \sigma^2) \in D_1^\lambda$. We choose $c = 1 - \lambda$ and $a = \mu - \lambda - \sigma^2/(1-\mu)$. In order to ensure $c_3 = 0$ (cf. (2.19)) we have to take

$$b = -c - 2a \equiv -2\mu - 1 + 3\lambda + \frac{2\sigma^2}{1-\mu}. \tag{2.22}$$

If $b \leq -\lambda$, then $P(t) \geq 0$ for all $-\lambda \leq t \leq 1-\lambda$. The inequality $b \leq -\lambda$ is equivalent to

$$\sigma^2 \leq (1-\mu)\left(\mu - 2\lambda + \frac{1}{2}\right) \equiv f_1 \Longleftrightarrow \left(\mu, \sigma^2\right) \in D_1^\lambda. \tag{2.23}$$

To complete the proof we note that the random variable $Y_* = X_* - \lambda$ with X_* defined by (2.2) assumes its values in the set $\{a, 1-\lambda\} \subset \{t : P(t) = 0\}$. To find the distribution of Y_* we use (2.17). Setting $a = \mu - \lambda - \sigma^2/(1-\mu)$ in (2.17) we obtain $q = 0$ and r, p as in (2.2).

(ii) Now $(\mu, \sigma^2) \in D_2^\lambda$ or, equivalently $\sigma^2 > f_1$ and $\sigma^2 > f_3$. Moreover, we can assume that $1/4 < \lambda < 3/4$ since only for such λ the region D_2^λ is nonempty. We choose $c = 1 - \lambda$ and $b = -\lambda$. Then $P(t) \geq 0$ for all $-\lambda \leq t \leq 1-\lambda$. In order to ensure $c_3 = 0$ (cf. (2.19)) we have to take

$$a = -\frac{b+c}{2} \equiv \lambda - \frac{1}{2}. \tag{2.24}$$

By our construction $\{t : P(t) = 0\} = \{-\lambda, a, 1-\lambda\}$. To find a distribution of Y_* supported by the set $\{-\lambda, a, 1-\lambda\}$ we use (2.17). It follows that $X_* = Y_* + \lambda$ has the distribution defined in (2.4).

(iii) We choose $c = -\lambda$ and $a = \mu - \lambda + \sigma^2/\mu$. In order to ensure $c_3 = 0$ (cf. (2.19)) we have to take

$$b = -c - 2a \equiv 3\lambda - 2\mu - \frac{2\sigma^2}{\mu}. \tag{2.25}$$

If $b \geq 1-\lambda$, then $P(t) \geq 0$ for all $-\lambda \leq t \leq 1-\lambda$. The inequality $b \geq 1-\lambda$ is equivalent to

$$\sigma^2 \leq \mu\left(2\lambda - \mu - \frac{1}{2}\right) \equiv f_3 \Longleftrightarrow \left(\mu, \sigma^2\right) \in D_3^{\lambda}. \tag{2.26}$$

To conclude the proof we notice that the random variable $Y_* = X_* - \lambda$ with X_* given by (2.7) assumes values from the set $\{-\lambda, a\} \subset \{t : P(t) = 0\}$. □

To prove Theorems 1.1 and 1.3 we apply Lemma 2.1 with $\lambda = \mu$. We provide the bounds of interest as Corollary 2.2. To prove the corollary it suffices to plug $\lambda = \mu$ in Lemma 2.1 and, using (2.2)–(2.7), to calculate $\mathbb{E}(X_* - \mu)^4$ explicitly. We omit related elementary however cumbersome calculations. The regions D_1, D_2, and D_3 are defined in (1.32).

Corollary 2.2. *Let a random variable $0 \leq X \leq 1$ have mean μ and variance σ^2. Then*

$$\mathbb{E}(X-\mu)^4 \leq \begin{cases} \sigma^6(1-\mu)^{-2} - \sigma^4 + \sigma^2(1-\mu)^2, & \text{if } (\mu,\sigma^2) \in D_1, \\ \mu(1-\mu)\left(\mu - \frac{1}{2}\right)^2 + \sigma^2\left(2\mu^2 - 2\mu + \frac{3}{4}\right), & \text{if } (\mu,\sigma^2) \in D_2, \\ \sigma^6\mu^{-2} - \sigma^4 + \sigma^2\mu^2, & \text{if } (\mu,\sigma^2) \in D_3. \end{cases} \tag{2.27}$$

Proposition 2.3. *Let $0 \leq X \leq 1$. Then, with probability 1, the sample variance satisfies $\widehat{\sigma}^2 \leq \gamma^2 + \sigma^2$ with γ^2 given by (1.6).*

Proof. Using the representation (1.3) of the sample variance as an U-statistic, it suffices to show that the function $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = \sum_{i \neq k,\ 1 \leq i,k \leq n} (x_i - x_k)^2, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{2.28}$$

in the domain

$$\mathfrak{D} = \{x \in \mathbb{R}^n : 0 \leq x_1 \leq 1, \ldots, 0 \leq x_n \leq 1\} \tag{2.29}$$

satisfies $f \leq 2n(n-1)(\gamma^2 + \sigma^2)$. The function f is convex. To see this, it suffices to check that f restricted to straight lines is convex. Any straight line can be represented as $L = \{x + th : t \in \mathbb{R}\}$

with some $x, h \in \mathbb{R}^n$. The convexity of f on L is equivalent to the convexity of the function $g(t) \stackrel{\text{def}}{=} f(x + th)$ of the real variable $t \in \mathbb{R}$. It is clear that the second derivative $g''(t) = 2f(h)$ is nonnegative since $f \geq 0$. Thus both g and f are convex.

Since both f and $\mathfrak{D}$ are convex, the function f attains its maximal value on the boundary of $\mathfrak{D}$. Moreover, the maximal value of f is attained on the set of extremal points of $\mathfrak{D}$. In our case the set of the extremal points is just the set of vertexes of the cube $\mathfrak{D}$. In other words, the maximal value of f is attained when each of $x_1, \ldots, x_n$ is either 0 or 1. Since f is a symmetric function, we can assume that the maximal value of f is attained when $x_1 = \cdots = x_m = 1$ and $x_{m+1} = \cdots = x_n = 0$ with some $m = 0, \ldots, n$. Using (2.28), the corresponding value of f is $2m(n-m)$. Maximizing with respect to m we get $f \leq n^2/2$, if n is even, and $f \leq (n^2-1)/2$, if n is odd, which we can rewrite as the desired inequality $f \leq 2n(n-1)\ (\gamma^2 + \sigma^2)$. □

3. Proofs

We use the following observation which in the case of an exponential function comes back to Hoeffding [1, Section 5]. Assume that we can represent a random variable, say T, as a weighted mixture of other random variables, say $T_1, \ldots, T_m$, so that

$$T = \alpha_1 T_1 + \cdots + \alpha_m T_m, \quad \alpha_1, \ldots, \alpha_m \geq 0,\ \alpha_1 + \cdots + \alpha_m = 1, \tag{3.1}$$

where α_s are nonrandom numbers. Let f be a convex function. Then, using Jensen's inequality $f(T) \leq \alpha_1 f(T_1) + \cdots + \alpha_m f(T_m)$, we obtain

$$\mathbb{E} f(T) \leq \alpha_1 \mathbb{E} f(T_1) + \cdots + \alpha_m \mathbb{E} f(T_m). \tag{3.2}$$

Moreover, if random variables $T_1, \ldots, T_m$ are identically distributed, then

$$\mathbb{E} f(T) \leq \mathbb{E} f(T_1). \tag{3.3}$$

One can specialize (3.3) for U-statistics of the second order. Let $u(x, y) = u(y, x)$ be a symmetric function of its arguments. For an i.i.d. sample $X_1, \ldots, X_n$ consider the U-statistic

$$U = \frac{1}{n(n-1)} \sum_{i \neq k,\ 1 \leq i, k \leq n} u(X_i, X_k). \tag{3.4}$$

Write

$$V = \frac{1}{k}\left(u(X_1, X_2) + u(X_3, X_4) + \cdots + u(X_{2k-1}, X_{2k})\right), \quad k = \left[\frac{n}{2}\right]. \tag{3.5}$$

Then (3.3) yields

$$\mathbb{E} f(U) \leq \mathbb{E} f(V) \tag{3.6}$$

for any convex function f. To see that (3.6) holds, let $\pi = (\pi(1), \ldots, \pi(n))$ be a permutation of $1, \ldots, n$. Define $V(\pi)$ as (3.5) replacing the sample $X_1, \ldots, X_n$ by its permutation $X_{\pi(1)}, \ldots, X_{\pi(n)}$. Then (see [1, Section 5])

$$U = \frac{1}{n!} \sum_{\pi} V(\pi), \tag{3.7}$$

which means that U allows a representation of type (3.1) with $m = n!$ and all $V(\pi)$ identically distributed, due to our symmetry and i.i.d. assumptions. Thus, (3.3) implies (3.6).

Using (1.3) we can write

$$T = \frac{1}{n(n-1)} \sum_{i \neq k,\ 1 \leq i,j \leq n} u(X_i, X_j) \tag{3.8}$$

with $u(x,y) = \sigma^2 - (x-y)^2/2$. By an application of (3.6) we derive

$$\mathbb{E} f(T) \leq \mathbb{E} f\left(\frac{Z_k}{k}\right), \quad \mathbb{E} f(-T) \leq \mathbb{E} f\left(-\frac{Z_k}{k}\right), \quad k = \left[\frac{n}{2}\right] \tag{3.9}$$

for any convex function f, where $Z_k = Y_1 + \cdots + Y_k$ is a sum of i.i.d. random variables such that

$$Y_1 \stackrel{\mathfrak{D}}{=} \sigma^2 - \frac{(X_1 - X_2)^2}{2}. \tag{3.10}$$

Consider the following three families of functions depending on parameters $t, h \in \mathbb{R}$:

$$f(y) = \frac{(y-h)_+}{(t-h)}, \quad t \in \mathbb{R},\ h < t, \tag{3.11}$$

$$f(y) = \frac{(y-h)_+^2}{(t-h)^2}, \quad t \in \mathbb{R},\ h < t, \tag{3.12}$$

$$f(y) = \exp\{h(y-t)\}, \quad t \in \mathbb{R},\ h > 0. \tag{3.13}$$

Any of functions f given by (3.11) dominates the indicator function $y \mapsto \mathbb{I}\{y \in [t, \infty)\}$ of the interval $[t, \infty)$. Therefore $\mathbb{P}\{T \geq t\} \leq \mathbb{E} f(T)$. Combining this inequality with (3.9), we get

$$\mathbb{P}\{T \geq t\} \leq \inf_h \mathbb{E}\, f\left(\frac{Z_k}{k}\right), \qquad \mathbb{P}\{T \leq -t\} \leq \inf_h \mathbb{E} f\left(-\frac{Z_k}{k}\right) \tag{3.14}$$

with Z_k being a sum of k i.i.d. random variables specified in (3.10). Depending on the choice of the family of functions f given by (3.11), the inf in (3.14) is taken over $h < t$ or $h > 0$, respectively.

Proposition 3.1. *One has*

$$\mathbb{E}(X_1 - X_2)^4 = 2\omega + 6\sigma^4. \tag{3.15}$$

If $0 \le X \le 1$, *then* $\omega = \mathbb{E}(X-\mu)^4 \le \sigma^2 - 3\sigma^4$.

Proof. Let us prove (3.15). Using the i.i.d. assumption, we have

$$\begin{aligned}\mathbb{E}(X_1 - X_2)^4 &= \mathbb{E}\big((X_1-\mu)+(\mu - X_2)\big)^4 \\ &= 2\mathbb{E}(X-\mu)^4 - 8\mathbb{E}(X_1-\mu)(X_2-\mu)^3 + 6\mathbb{E}(X_1-\mu)^2(X_2-\mu)^2 \\ &= 2\omega + 6\sigma^4.\end{aligned} \tag{3.16}$$

Let us prove that $\omega \le \sigma^2 - 3\sigma^4$. If $0 \le X \le 1$, then $(X_1 - X_2)^2 \le 1$. Using (3.15) we have

$$2\omega + 6\sigma^4 = \mathbb{E}(X_1 - X_2)^4 \le \mathbb{E}(X_1 - X_2)^2 = 2\sigma^2, \tag{3.17}$$

which yields the desired bound for ω. □

Proposition 3.2. *Let* Y *be a bounded random variable such that* $a \le Y \le b$ *with some nonrandom* $a, b \in \mathbb{R}$. *Then for any convex function* $g : [a,b] \to \mathbb{R}$ *one has*

$$\mathbb{E}g(Y) \le \mathbb{E}g(\varepsilon), \tag{3.18}$$

where ε *is a Bernoulli random variable such that* $\mathbb{E}Y = \mathbb{E}\varepsilon$ *and* $\mathbb{P}\{\varepsilon = a\} + \mathbb{P}\{\varepsilon = b\} = 1$.

If $Y \le b$ *for some* $b > 0$, *and* $\mathbb{E}Y = 0$, $\mathbb{E}Y^2 \le r^2$, *then* (3.18) *holds with*

$$g(y) = (y-h)_+^2, \quad h \in \mathbb{R}, \qquad g(y) = \exp\{hy\}, \quad h \ge 0, \tag{3.19}$$

and a Bernoulli random variable ε *such that* $\mathbb{E}\varepsilon = 0$, $\operatorname{var}\varepsilon = r^2$,

$$p_r \overset{\text{def}}{=} \mathbb{P}\{\varepsilon = b\} = \frac{r^2}{b^2 + r^2}, \qquad q_r \overset{\text{def}}{=} \mathbb{P}\left\{\varepsilon = -\frac{r^2}{b}\right\} = \frac{b^2}{b^2 + r^2}, \tag{3.20}$$

$p_r + q_r = 1$.

Proof. See [2, Lemmas 4.3 and 4.4]. □

Proof of Theorem 1.1. The proof is based on a combination of Hoeffding's observation (3.6) using the representation (3.8) of T as a U-statistic, of Chebyshev's inequality involving exponential functions, and of Proposition 3.2. Let us provide more details. We have to prove (1.22) and (1.24).

Let us prove (1.22). We apply (3.14) with the family (3.13) of exponential functions f. We get

$$\mathbb{P}\{T \geq t\} \leq \inf_{h>0} \exp\{-ht\} \mathbb{E} \exp\left\{ \frac{hZ_k}{k} \right\}. \tag{3.21}$$

By (3.10), the sum $Z_k = Y_1 + \cdots + Y_k$ is a sum of k copies of a random variable, say Y, such that

$$Y = \sigma^2 - \frac{(X_1 - X_2)^2}{2}. \tag{3.22}$$

We note that

$$Y \leq \sigma^2, \quad \mathbb{E}Y = 0, \quad \mathbb{E}Y^2 = \frac{\omega + \sigma^4}{2} \leq \frac{\omega_0 + \sigma^4}{2}. \tag{3.23}$$

Indeed, the first two relations in (3.23) are obvious; the third one is implied by $\omega \leq \omega_0$,

$$\mathbb{E}Y^2 = \frac{\mathbb{E}(X_1 - X_2)^4}{4} - \sigma^4, \tag{3.24}$$

and $\mathbb{E}(X_1 - X_2)^4 = 2\omega + 6\sigma^4$; see Proposition 3.1.

Let $\mathcal{M}$ stand for the class of random variables Y satisfying (3.23). Taking into account (3.21), to prove (1.22) it suffices to check that

$$J \stackrel{\text{def}}{=} \inf_{h>0} \exp\{-ht\} \sup_{Y_1,\ldots,Y_k \in \mathcal{M}} \mathbb{E} \exp\left\{ \frac{hZ_k}{k} \right\} = H^k\left(\frac{t}{\sigma^2}; p \right), \tag{3.25}$$

where Z_k is a sum of k independent copies $Y_1, \ldots, Y_k$ of Y. It is clear that the left-hand side of (3.25) is an increasing function of ω_0. To prove (3.25), we apply Proposition 3.2. Conditioning k times on all random variables except one, we can replace all random variables $Y_1, \ldots, Y_k$ by Bernoulli ones. To find the distribution of the Bernoulli random variables we use (3.23). We get

$$\sup_{Y \in \mathcal{M}} \mathbb{E} \exp\left\{ \frac{hZ_k}{k} \right\} = \mathbb{E} \exp\left\{ \frac{hS_k}{k} \right\}, \tag{3.26}$$

where $S_k = \varepsilon_1 + \cdots + \varepsilon_k$ is a sum of k independent copies of a Bernoulli random variable, say ε, such that $\mathbb{E}\varepsilon = 0$ and $\mathbb{P}\{\varepsilon = \sigma^2\} = p$ with p as in (1.23), that is, $p = (\sigma^4 + \omega_0)/(3\sigma^4 + \omega_0)$. Note that in (3.26) we have the equality since $\varepsilon \in \mathcal{M}$.

Using (3.26) we have

$$\begin{aligned} J &= \inf_{h>0} \exp\{-ht\}\mathbb{E}\exp\left\{\frac{hS_k}{k}\right\} \\ &= \left(\inf_{h>0} \exp\left\{-\frac{ht}{k}\right\}\mathbb{E}\exp\left\{\frac{h\varepsilon}{k}\right\}\right)^k \\ &= \left(\inf_{h>0} \exp\left\{-\frac{ht}{\sigma^2}\right\}\mathbb{E}\exp\left\{\frac{h\varepsilon}{\sigma^2}\right\}\right)^k \\ &= H^k\left(\frac{t}{\sigma^2};p\right). \end{aligned} \tag{3.27}$$

To see that the third equality in (3.27) holds, it suffices to change the variable h by kh/σ^2. The fourth equality holds by definition (1.13) of the Hoeffding function since ε/σ^2 is a Bernoulli random variable with mean zero and such that $\mathbb{P}\{\varepsilon/\sigma^2 = 1\} = p$. The relation (3.27) proves (3.25) and (1.22).

A proof of (1.24) repeats the proof of (1.22) replacing everywhere T and Y by $-T$ and $-Y$, respectively. The inequality $Y \le \sigma^2$ in (3.23) has to be replaced by $-Y \le 1/2 - \sigma^2$, which holds due to our assumption $0 \le X \le 1$. Respectively, the probability p now is given by (1.25). □

Proof of (1.19). The bound is an obvious corollary of Theorem 1.1 since by Proposition 3.1 we have $\omega \le \sigma^2 - 3\sigma^4$, and therefore we can choose $\omega_0 = \sigma^2 - 3\sigma^4$. Setting this value of ω_0 into (1.22), we obtain (1.19). □

Proof of (1.26) *and* (1.27). To prove (1.26), we set $\omega_0 = \sigma^2 - 3\sigma^4$ in (1.24). Such choice of ω_0 is justified in the proof of (1.19).

To prove (1.27) we use (1.26). We have to prove that

$$H\left(\frac{2t}{1-2\sigma^2};2\sigma^2\right) \le H\left(2t;\frac{2\sigma^2}{1+2\sigma^2}\right), \tag{3.28}$$

and that the right-hand side of (3.28) is an increasing function of σ^2. By the definition of the Hoeffding function we have

$$\begin{aligned} H\left(\frac{2t}{1-2\sigma^2};2\sigma^2\right) &= \inf_{h>0} \exp\left\{-\frac{2ht}{1-2\sigma^2}\right\}\mathbb{E}\exp\{h\delta\} \\ &= \inf_{h>0} \exp\{-2ht\}\mathbb{E}\exp\left\{h\left(1-2\sigma^2\right)\delta\right\}, \end{aligned} \tag{3.29}$$

where δ is a Bernoulli random variable such that $\mathbb{P}\{\delta = 1\} = 2\sigma^2$ and $\mathbb{E}\delta = 0$. It is easy to check that δ assumes as well the value $-2\sigma^2/(1-2\sigma^2)$ with probability $1-2\sigma^2$. Hence $-2\sigma^2/(1-2\sigma^2) \le \delta \le 1$. Therefore $-2\sigma^2 \le (1-2\sigma^2)\delta \le 1-2\sigma^2$, and we can write

$$\mathbb{E}\exp\left\{h\left(1-2\sigma^2\right)\delta\right\} \le \sup_{W\in\mathcal{M}} \mathbb{E}\exp\{hW\}, \tag{3.30}$$

where $\mathcal{M}$ is the class of random variables W such that $\mathbb{E}W = 0$ and $-2\sigma^2 \le W \le 1$. Combining (3.29) and (3.30) we obtain

$$H\left(\frac{2t}{1-2\sigma^2}; 2\sigma^2\right) \le \inf_{h>0} \exp\{-2ht\} \sup_{W\in\mathcal{M}} \mathbb{E}\exp\{hW\}. \tag{3.31}$$

The definition of the latter sup in (3.31) shows that the right-hand side of (3.31) is an increasing function of σ^2. To conclude the proof of (1.27) we have to check that the right-hand sides of (3.28) and (3.31) are equal. Using (3.18) of Proposition 3.2, we get $\mathbb{E}\exp\{hW\} \le \mathbb{E}\exp\{h\varepsilon\}$, where ε is a mean zero Bernoulli random variable assuming the values $-2\sigma^2$ and 1 with positive probabilities such that $\mathbb{P}\{\varepsilon = 1\} = 2\sigma^2/(1+2\sigma^2)$. Since $\varepsilon \in \mathcal{M}$, we have

$$\sup_{W\in\mathcal{M}} \mathbb{E}\exp\{hW\} = \mathbb{E}\exp\{h\varepsilon\}. \tag{3.32}$$

Using the definition of the Hoeffding function we see that the right-hand sides of (3.28) and (3.31) are equal. □

Proof of Theorem 1.3. We use Theorem 1.1. In bounds of this theorem we substitute the value of ω_0 being the right-hand side of (2.27), where a bound of type $\omega \le \omega_0$ is given. We omit related elementary analytical manipulations. □

Proof of the Asymptotic Relations (1.7) *and* (1.8). To describe the limiting behavior of T we use Hoeffding's decomposition. We can write

$$\frac{n(n-1)}{2\sigma^2}T = (n-1)\sum_{1\le i\le n} u_1(X_i) + \sum_{1\le i<k\le n} u_2(X_i, X_k) \tag{3.33}$$

with kernels u_1 and u_2 such that

$$u_1(x) = \frac{\sigma^2 - (x-\mu)^2}{2\sigma^2}, \qquad u_2(x) = \frac{(x-\mu)(y-\mu)}{\sigma^2}. \tag{3.34}$$

To derive (3.33), use the representation of T as a U-statistic (3.8). The kernel functions u_1 and u_2 are degenerated, that is, $\mathbb{E}u_1(X) = 0$ and $\mathbb{E}u_2(X, x) = 0$ for all $x \in \mathbb{R}$. Therefore

$$\operatorname{var}\left(\frac{n(n-1)}{2\sigma^2}T\right) = n(n-1)^2 \operatorname{var} u_1 + \frac{n(n-1)}{2}\operatorname{var} u_2 \tag{3.35}$$

with

$$\operatorname{var} u_1 = \frac{\omega - \sigma^4}{4\sigma^4}, \quad \operatorname{var} u_2 = 1, \quad \omega = \mathbb{E}(X-\mu)^4. \tag{3.36}$$

It follows that in cases where $\omega > \sigma^4$ the statistic T is asymptotically normal:

$$\frac{\sqrt{n}T}{\sqrt{\omega - \sigma^4}} \longrightarrow \eta, \quad \text{as } n \longrightarrow \infty, \tag{3.37}$$

where η is a standard normal random variable. It is easy to see that $\omega = \sigma^4$ if and only if X is a Bernoulli random variable symmetric around its mean. In this special case we have $u_1(X) \equiv 0$, and (3.33) turns to

$$\frac{n(n-1)T}{\sigma^2} \overset{\mathfrak{D}}{=} (\varepsilon_1 + \cdots + \varepsilon_n)^2 - n, \tag{3.38}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. Rademacher random variables. It follows that

$$\omega = \sigma^4 \Longrightarrow \frac{(n-1)T}{\sigma^2} \longrightarrow \eta^2 - 1, \quad \text{as } n \longrightarrow \infty, \tag{3.39}$$

which completes the proof of (1.7) and (1.8). □

Acknowledgment

Figure 1 was produced by N. Kalosha. The authors thank him for the help. The research was supported by the Lithuanian State Science and Studies Foundation, Grant no. T-15/07.

References

[1] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.
[2] V. Bentkus and M. van Zuijlen, "On conservative confidence intervals," *Lithuanian Mathematical Journal*, vol. 43, no. 2, pp. 141–160, 2003.
[3] M. Talagrand, "The missing factor in Hoeffding's inequalities," *Annales de l'Institut Henri Poincaré B*, vol. 31, no. 4, pp. 689–702, 1995.
[4] I. Pinelis, "Optimal tail comparison based on comparison of moments," in *High Dimensional Probability (Oberwolfach, 1996)*, vol. 43 of *Progress in Probability*, pp. 297–314, Birkhäuser, Basel, Switzerland, 1998.
[5] I. Pinelis, "Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities," in *Advances in Stochastic Inequalities (Atlanta, Ga, 1997)*, vol. 234 of *Contemporary Mathematics*, pp. 149–168, American Mathematical Society, Providence, RI, USA, 1999.
[6] V. Bentkus, "A remark on the inequalities of Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand," *Lithuanian Mathematical Journal*, vol. 42, no. 3, pp. 262–269, 2002.
[7] V. Bentkus, "On Hoeffding's inequalities," *The Annals of Probability*, vol. 32, no. 2, pp. 1650–1673, 2004.
[8] V. Bentkus, G. D. C. Geuze, and M. van Zuijlen, "Trinomial laws dominating conditionally symmetric martingales," Tech. Rep. 0514, Department of Mathematics, Radboud University Nijmegen, 2005.
[9] V. Bentkus, N. Kalosha, and M. van Zuijlen, "On domination of tail probabilities of (super)martingales: explicit bounds," *Lithuanian Mathematical Journal*, vol. 46, no. 1, pp. 3–54, 2006.