PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 30, Fasc. 2 (2010), pp. 339–351

INEQUALITIES FOR QUANTILES OF THE CHI-SQUARE DISTRIBUTION

BY

TADEUSZ INGLOT (WROCŁAW)

Abstract. We obtain a new sharp lower estimate for tails of the central chi-square distribution. Using it we prove quite accurate lower bounds for the chi-square quantiles covering the case of increasing number of degrees of freedom and simultaneously tending to zero tail probabilities. In the case of small tail probabilities we also provide upper bounds for these quantiles which are close enough to the lower ones. As a byproduct we propose a simple approximation formula which is easy to calculate for the chi-square quantiles. It is expressed explicitly in terms of tail probabilities and a number of degrees of freedom.

2000 AMS Mathematics Subject Classification: Primary: 62E17; Secondary: 60E15, 62E15, 62Q05, 65C60.

Key words and phrases: Chi-square quantiles, normal quantiles, lower and upper bounds, tails of chi-square distribution, Wilson–Hilferty formula.

1. INTRODUCTION

Approximation formulae for the chi-square distribution quantiles have been investigated in numerous papers beginning from Fisher [3] and Wilson and Hilferty [9]. Nowadays, very accurate approximation formulae are available (see e.g. Zar [10] and Johnson et al. [6], [7] or Ittrich et al. [5] and references therein). On the contrary, inequalities for the central chi-square quantiles rarely appear in the literature although they play an important role in some statistical considerations. Laurent and Massart [8] gave an exponential inequality for tails of the noncentral chi-square distrubution and used it to determine risk bounds for penalized estimator of the squared norm of a mean in a Gaussian linear model. This inequality is equivalent to some global upper bound for quantiles covering all values of parameters involved. Brain and Mi [2] proved some upper and lower bounds which are expressed solely in terms of a number of degrees of freedom k and applied them to an interval estimation problem. Inglot and Ledwina [4] obtained lower bounds depending both on k and tail probabilities α and, employing them, described an asymptotic behaviour of the quantiles when k increases and simultaneously α tends to zero. Such considerations were necessary to study an asymptotic optimality of some adaptive test proposed by Baraud et al. [1].

In the present note we prove a new sharp lower estimate for tails of the central chi-square distribution. Next, using it, results of Inglot and Ledwina [4] are essentially improved. In effect, we provide accurate upper and lower bounds for the chi-square quantiles for small α. Their form suggests to propose a new simple approximation formula for the chi-square quantiles which for typical k between 3 and 100 and typical α between 0.1 and 0.0001 gives comparable relative errors as the celebrated Wilson–Hilferty formula. Being expressed explicitly in terms of k and α it is easy for hand calculations which may be regarded as its additional advantage. For the sake of completeness, we provide also some lower and upper bound for the normal quantiles.

2. BOUNDS FOR THE NORMAL QUANTILES

In this section we give some bounds for quantiles of the standard normal distribution. First, recall the well-known inequality for tails of this distribution which we shall use to derive our result. Namely, let $\Phi(x)$ be the standard normal cdf. Then for every $u > 0$ we have

$$\frac{u^2}{1+u^2}h(u) \leqslant 1 - \Phi(u) \leqslant h(u), \tag{2.1}$$

where $h(u) = \exp\{-\frac{1}{2}(u^2 + 2\log u + \log 2\pi)\}$. Note that the upper bound equals $\phi(u)/u$ and the lower bound equals $u\phi(u)/(1+u^2)$, where $\phi(u)$ is the standard normal density.

Now, denote by z_α the $1-\alpha$ quantile of the standard normal distribution, i.e. defined by the equation $\Phi(z_\alpha) = 1 - \alpha$.

Our result is intended especially for small α, say, $\alpha \leqslant 0.1$. Then, applying (2.1), we obtain the following estimate for z_α:

THEOREM 2.1. *For every* $0 < \alpha \leqslant 0.1$ *we have*

$$\sqrt{2\log(1/\alpha)} - \frac{\log\big(4\log(1/\alpha)\big) + 2}{2\sqrt{2\log(1/\alpha)}} \leqslant z_\alpha \leqslant \sqrt{2\log(1/\alpha)} - \frac{\log\big(2\log(1/\alpha)\big) + 3/2}{2\sqrt{2\log(1/\alpha)}}. \tag{2.2}$$

The proof of Theorem 2.1 is provided in Section 6.

3. TAILS OF THE CHI-SQUARE DISTRIBUTION

Let χ_k^2 denote a random variable with central chi-square distribution with k degrees of freedom. For various applications explicit lower and upper bounds for tails of χ_k^2 are important. Some estimates are well known. In Lemma 1 of Inglot and Ledwina [4], the following version of such an estimate was proved:

$$\frac{1}{2}\mathcal{E}_k(u) \leqslant P(\chi_k^2 \geqslant u) \leqslant \frac{1}{\sqrt{\pi}}\frac{u}{u-k+2}\mathcal{E}_k(u), \qquad k \geqslant 2,\ u > k-2, \tag{3.1}$$

where $\mathcal{E}_k(u) = \exp\left\{-\frac{1}{2}\big(u-k-(k-2)\log(u/k)+\log k\big)\right\}$. It follows from the proof of (3.1) that the lower bound holds true for all $u > 0$ and any $k \geqslant 2$. This observation will be used in (6.7) below. The upper bound in (3.1) seems to be accurate enough while the lower bound for u not much greater than k is far from being precise and differs from the upper one significantly. In order to manage this problem we propose an essentially better lower bound than that given in (3.1).

PROPOSITION 3.1. *For all $k \geqslant 2$ and all $u \geqslant k-1$ we have*

$$P(\chi_k^2 \geqslant u) \geqslant \frac{1-e^{-2}}{2}\frac{u}{u-k+2\sqrt{k}}\mathcal{E}_k(u). \tag{3.2}$$

Proposition 3.1 is proved in Section 6.

4. UPPER BOUNDS FOR THE CHI-SQUARE QUANTILES

From now on denote by $u(\alpha, k)$ the quantile of order $1-\alpha$, $\alpha \in (0,1)$, of the central chi-square distribution with k degrees of freedom, i.e. satisfying the relation

$$P\left(\chi_k^2 \geqslant u(\alpha,k)\right) = \alpha. \tag{4.1}$$

A global upper bound for quantiles of the noncentral chi-square distribution follows from Lemma 1 of Laurent and Massart [8]. In a special case of the central chi-square distribution it may be stated as follows:

THEOREM A (Laurent and Massart [8]). *For every $k \geqslant 1$ and every $\alpha \in (0,1)$ we have*

$$u(\alpha,k) \leqslant k + 2\log(1/\alpha) + 2\sqrt{k\log(1/\alpha)}. \tag{4.2}$$

The relation (4.2) is a simple consequence of the upper bound in (3.1). That is why we reprove Theorem A in Section 6.

As one can expect, (4.2) being true for the whole range of k and α is far from being precise. In particular, the last term on the right-hand side takes much too large values for very small α. More precisely, the following fact takes place.

THEOREM 4.1. *For any constant* $C \in (0,2)$ *there exists* $a = a(C) > 0$ *such that for every* $k \geqslant 1$ *and every* $\alpha \leqslant e^{-ak}$ *we have*

$$u(\alpha, k) \leqslant k + 2\log(1/\alpha) + C\sqrt{k\log(1/\alpha)}. \tag{4.3}$$

The proof of Theorem 4.1 is given in Section 6 and is a modification of our proof of Theorem A. To see to what range of α it applies in the assumption take, for example, $C = 1$ and $C = 1/4$. Then from (6.11) it follows that $a(1) \approx 10.33$ and $a(1/4) \approx 900$.

Theorem 4.1 suggests that for very small α the third term in (4.3) being of the form $\sqrt{k\log(1/\alpha)}$ has too large order. An upper bound in which this term has more proper order is stated in the next theorem and proved in Section 6.

THEOREM 4.2. *For every* $k \geqslant 2$, $0 < d < 1$ *and every* α,

$$\alpha \leqslant \exp\{-\exp\left([(k-2)/2\sqrt{k}]^{1/d}\right)\},$$

we have

$$u(\alpha, k) \leqslant k + 2\log(1/\alpha) + 2\sqrt{k}\log^{1+d}\log(1/\alpha). \tag{4.4}$$

The above theorem shows that a proper order of the third term in expansion of the quantiles for very small α seems to be $\sqrt{k}\log\log(1/\alpha)$. We shall see in the next section that, in fact, it is the case.

5. LOWER BOUNDS FOR THE CHI-SQUARE QUANTILES

In some statistical considerations lower bounds for the chi-square quantiles seem to be as useful as the upper ones. For example, both lower and upper bounds were used by Brain and Mi [2] to prove some properties of confidence bounds for the maximum likelihood estimators. However, it turns out that it is hard to find in the literature sufficiently accurate lower bounds for $u(\alpha, k)$ in the case when α tends to zero and simultaneously k increases. As was said previously, some results were obtained, among others, in Lemma 3 of Inglot and Ledwina [4]. A global lower bound which may be considered as a counterpart of the Laurent and Massart upper bound was proposed in (4) of Lemma 3 in [4]. Using (3.2) it can be significantly improved. The corresponding result is stated in the next proposition.

PROPOSITION 5.1. *For every* $k \geqslant 2$ *and every* $\alpha \leqslant 0.17$ *we have*

$$u(\alpha, k) \geqslant k + 2\log(1/\alpha) - 5/2. \tag{5.1}$$

A simple proof of Proposition 5.1 is given in Section 6. A similar proof shows that at the cost of enlarging k to $k \geqslant 9$ and diminishing α to $\alpha \leqslant 1/17$ one can remove the constant $5/2$ in (5.1) to obtain $u(\alpha, k) \geqslant k + 2\log(1/\alpha)$. However, both

(5.1) as well as (5) and (6) of Lemma 3 in [4] leave a wide gap to the upper bounds (4.4) and (4.2). So, it seems desirable to look for better bounds by an application of (3.2). Note that for $u \geqslant k + 2\log(1/\alpha)$ the estimate (3.2) can be better than (3.1) only if k is not too small. Below we give for such k's a quite accurate lower bound much better than that in (5) of Lemma 3 in [4] and which corresponds to the logarithmic order term in expansion in (4.4). Its proof needs somewhat delicate numerical considerations and is given in Section 6.

THEOREM 5.1. *For every* $k \geqslant 32$ *and every* $\alpha \leqslant e^{-k/8}$ (*or any* $k \geqslant 18$ *and every* $\alpha \leqslant e^{-k/3}$) *we have*

$$u(\alpha, k) \geqslant k + 2\log(1/\alpha) + \sqrt{k}\log\log(1/\alpha). \tag{5.2}$$

If α is not very close to 0, then the order of the third term in (5.2) can be enlarged to that appearing in the upper bounds (4.2) and (4.3). A result of such a type is given in the next theorem. It improves (6) of Lemma 3 in [4].

THEOREM 5.2. *For every* $k \geqslant 17$ *and every* $\alpha \in [e^{-560k}, 1/17]$ *we have*

$$u(\alpha, k) \geqslant k + 2\log(1/\alpha) + \frac{1}{4}\sqrt{k\log(1/\alpha)}. \tag{5.3}$$

It is worth to mention that, in fact, (5.2), (5.3) as well as (4.4) remain true for much wider ranges of α and k than we are able to prove. For example, numerical calculations show that (5.2) holds true for all $k \geqslant 18$ and $\alpha \leqslant 0.05$ or for all $k \geqslant 28$ and $\alpha \leqslant 0.2$.

6. PROOFS

Proof of Theorem 2.1. Let us put $t = \log(1/\alpha)$. Then $t \geqslant \log 10$. Let us write $z^* = \sqrt{2t} - (\log 4t + 2)/2\sqrt{2t}$ and $z^{**} = \sqrt{2t} - (\log 2t + 3/2)/2\sqrt{2t}$. To prove the right-hand side of the estimate (2.2) it is enough, due to (2.1), to show that $h(z^{**}) \leqslant \alpha$ or, equivalently,

$$(z^{**})^2 + 2\log z^{**} + \log 2\pi \geqslant 2t. \tag{6.1}$$

Inserting the form of z^{**} one can reduce (6.1) to the inequality

$$\frac{\log 2t + 3/2}{2} v + 2\log(1-v) + \log 2\pi \geqslant \frac{3}{2},$$

where we have defined $v = v(t) = (\log 2t + 3/2)/4t$. Since v is decreasing with respect to t, it takes values in $(0, 1/3)$. As $\log 2t > 3/2$, it is enough to show

$$3v + 4\log(1-v) + 2\log 2\pi \geqslant 3, \quad v \in \left(0, \frac{1}{3}\right). \tag{6.2}$$

The left-hand side of (6.2) is decreasing with respect to v and its minimal value is attained at $v = 1/3$, so it is greater than $1 + 2\log(8\pi/9)$ which, in turn, is greater than 3.

To prove the left-hand side of (2.2) it is enough, again by (2.1), to check $h(z^*)(z^*)^2/\big(1+(z^*)^2\big) \geqslant \alpha$. Putting $u = u(t) = (\log 4t + 2)/4t$ and repeating similar calculations as above we reduce this inequality to an equivalent form

$$(6.3)\qquad 2tu^2 + 2\log\left(1 - u + \frac{1}{2t(1-u)}\right) + \log\pi \leqslant 2.$$

Applying the inequality $\log(1+y) \leqslant y$ to the second term in (6.3) it is enough to prove

$$(6.4)\qquad \frac{\log^2 4t}{8t} + \frac{-1 + 2/(1-u)}{2t} + \log\pi \leqslant 2 \quad \text{for } t \geqslant \log 10.$$

Since u is decreasing with respect to t in $(\log 10, \infty)$, the left-hand side of (6.4) is also decreasing with respect to t and its largest value is attained at $t = \log 10$, being obviously less than 2. This completes the proof. ■

Proof of Proposition 3.1. For $k \geqslant 4$ and $u > 0$ integration by parts gives

$$(6.5)\qquad \int_u^\infty x^{(k-2)/2} e^{-x/2} dx = 2u^{(k-2)/2} e^{-u/2} + (k-2)\int_u^\infty x^{(k-4)/2} e^{-x/2} dx.$$

From the proof of Lemma 1 in [4] we obtain $c(k) = 2(k/2e)^{k/2}/\big(\sqrt{k}\,\Gamma(k/2)\big) > 1/2$ for all $k \geqslant 2$, which together with (6.5) imply that for $k \geqslant 4$ and $u > 0$ the tail probabilities of the chi-square distribution satisfy a recurrence formula

$$(6.6)\qquad P(\chi_k^2 \geqslant u) \geqslant \frac{1}{2}\mathcal{E}_k(u) + P(\chi_{k-2}^2 \geqslant u).$$

Let us put $r_k = \lfloor\sqrt{k}\rfloor$, where $\lfloor a \rfloor$ is the integer part of a number a. If $k \geqslant 6$, then iteration of (6.6) r_k times yields

(6.7)
$$\begin{aligned} P(\chi_k^2 \geqslant u) &\geqslant \frac{1}{2}\mathcal{E}_k(u) + \frac{1}{2}\mathcal{E}_{k-2}(u) + \ldots + \frac{1}{2}\mathcal{E}_{k-2r_k+2}(u) + P(\chi_{k-2r_k}^2 \geqslant u) \\ &\geqslant \frac{1}{2}\mathcal{E}_k(u) + \ldots + \frac{1}{2}\mathcal{E}_{k-2r_k}(u) \end{aligned}$$

for $u > 0$, where the last inequality follows from (3.1).

Since $[k/(k-2)]^{(k-1)/2} \geqslant e$ for all $k \geqslant 3$ (insert $x = 2/(k-2)$ into the inequality $(x+2)\log(x+1) \geqslant 2x$ which holds for $x \geqslant 0$), we infer that $\mathcal{E}_{k-2}(u) \geqslant [(k-2)/u]\mathcal{E}_k(u)$ for $k \geqslant 3$ and $u > 0$. Consequently, for $k \geqslant 6$ and $u > 0$

$$(6.8)\qquad \begin{aligned} P(\chi_k^2 \geqslant u) &\geqslant \frac{1}{2}\mathcal{E}_k(u)\left[1 + \frac{k-2}{u} + \ldots + \frac{(k-2)\ldots(k-2r_k)}{u^{r_k}}\right] \\ &\geqslant \frac{1}{2}\mathcal{E}_k(u)\left[1 + \frac{k-2\sqrt{k}}{u} + \ldots + \left(\frac{k-2\sqrt{k}}{u}\right)^{r_k}\right] \\ &= \frac{1}{2}\mathcal{E}_k(u)\frac{u}{u-k+2\sqrt{k}}\left[1 - \left(\frac{k-2\sqrt{k}}{u}\right)^{r_k+1}\right]. \end{aligned}$$

It is easy to see that for $u \geqslant k-2$ we have

$$\left(\frac{k-2\sqrt{k}}{u}\right)^{r_k+1} \leqslant \left(\frac{k-2\sqrt{k}}{k-2}\right)^{r_k+1} \leqslant e^{-2},$$

which together with (6.8) proves (3.2) for $k \geqslant 6$. For $k = 2, 3, 4$ and 5 the right-hand side of (3.2) is for $u \geqslant k-1$ smaller than $\mathcal{E}_k(u)/2$ and our result follows immediately from (3.1). ■

REMARK 6.1. More careful considerations in (6.8) lead to a slightly stronger estimate than (3.2), which has the form

$$P(\chi_k^2 \geqslant u) \geqslant \frac{1}{2}\frac{u - e^{-2}k + 2\sqrt{k} - 2}{u - k + 2\sqrt{k}}\mathcal{E}_k(u)$$

and which holds true for all $k \geqslant 6$ and $u \geqslant k$. Using this estimate, we can further improve (5.2) and (5.3). We omit details.

In all proofs below we shall still use the notation $t = \log(1/\alpha)$ for $\alpha \in (0, 1)$. Obviously, the relation $u(\alpha, k) \leqslant u^{**}$ holds if and only if $P(\chi_k^2 \geqslant u^{**}) \leqslant \alpha$. Hence, to get this inequality for $k \geqslant 2$ it is sufficient to show, due to the upper bound in (3.1), that

$$(6.9)\qquad \frac{u^{**}}{u^{**} - k + 2}\mathcal{E}_k(u^{**}) \leqslant \sqrt{\pi}\alpha.$$

P r o o f o f T h e o r e m A. For $k = 1$, (4.2) follows immediately from (3.7) of Lemma 3 in [4] or from our Theorem 2.1 after some easy calculations. For $k \geqslant 2$ we need to check (6.9) with $u^{**} = k + 2t + 2\sqrt{kt}$. Inserting this form of u^{**} into (6.9), taking logarithms of both sides and rearranging we get an equivalent form of (6.9):

$$2\sqrt{kt} - k\log\left(1 + \frac{2t}{k} + 2\sqrt{\frac{t}{k}}\right) + 2\log\left(\frac{2}{\sqrt{k}} + \frac{2t}{\sqrt{k}} + 2\sqrt{t}\right) + \log\pi \geqslant 0.$$

Putting $t = kv$ we obtain finally

$$kh(v) + 2\log\left(\frac{2}{\sqrt{k}} + 2\sqrt{kv} + 2\sqrt{t}\right) + \log\pi \geqslant 0, \tag{6.10}$$

where $h(v) = 2\sqrt{v} - \log(1 + 2v + 2\sqrt{v})$. Since $h(v)$ is increasing on $(0, \infty)$ and $h(0) = 0$, the first term in (6.10) is positive. The sum of the two last terms in (6.10) is greater than $\log(4\pi t)$ which, in turn, is positive for $t > 1/(4\pi)$, i.e. for $\alpha < e^{-1/(4\pi)} \simeq 0.92$. But, obviously, for $\alpha > 0.5$ quantiles $u(\alpha, k)$ are smaller than k, thus satisfy (4.2). This completes the proof. ■

Proof of Theorem 4.1. For $y > 0$ let us define the function $g(y) = (e^y - 1 - y)/y^2$. Then g is increasing on $(0, \infty)$ and takes values in $(1/2, \infty)$. Put

$$a = a(C) = [g^{-1}(2/C^2)/C]^2. \tag{6.11}$$

Inserting $u^{**} = k + 2t + C\sqrt{kt}$ into (6.9) and repeating the same calculations as in the preceding proof we see that it is enough to show that

$$k\left(C\sqrt{v} - \log(1 + 2v + C\sqrt{v})\right) + 2\log\left(\frac{2}{\sqrt{k}} + \frac{2t}{\sqrt{k}} + C\sqrt{t}\right) + \log\pi \geqslant 0 \tag{6.12}$$

for $t \geqslant ak$ or, equivalently, $v \geqslant a$. The relation $v \geqslant a$ is equivalent to $g(C\sqrt{v}) \geqslant 2/C^2$, which in turn means that the first term in (6.12) is nonnegative. Obviously, $g(\sqrt{2}) < 1$. Consequently, for $C \leqslant \sqrt{2}$ we have $a > 1$ and for $t \geqslant ak > k$ the second term in (6.12) is positive, which proves (6.12) in this case. If $C > \sqrt{2}$ and $t > 1/2$, then the second term in (6.12) is positive. Finally, for $C > \sqrt{2}$ and $t \leqslant 1/2$, i.e. for $\alpha \geqslant e^{-1/2} > 1/2$, the inequality (4.3) is trivially satisfied, as was seen in the previous proof. This concludes the proof. ■

Proof of Theorem 4.2. For $k = 2$ the relation (4.3) is obvious. So, assume $k \geqslant 3$. Applying (6.9) to $u^{**} = k + 2t + 2\sqrt{k}\log^{1+d} t$, repeating the same calculations as in the previous proofs and setting $\log t = u$ we see that it is enough to check that

$$\psi_k(u) + 2\log(1 + e^{-u} + \sqrt{k}u^{1+d}e^{-u}) + \log\frac{4\pi}{k} \geqslant 0 \tag{6.13}$$

for $k \geqslant 3,\ 0 < d < 1$ and $u \geqslant u_0(k, d) = \left((k-2)/2\sqrt{k}\right)^{1/d}$, where

$$\psi_k(u) = 2\sqrt{k}u^{1+d} + 2u - k\log\left(1 + \frac{2}{k}e^u + \frac{2}{\sqrt{k}}u^{1+d}\right).$$

A standard calculation shows that for every $k \geqslant 3$ and $0 < d < 1$ the function $\psi_k(u)$ is increasing on $(u_0/2, \infty)$.

For $k \geqslant 8$ we omit the second term in (6.13) and it remains to show that

$$\psi_k(u_0) + \log\frac{4\pi}{k} = -k\log\left(\frac{2}{k} + e^{-u_0} + \frac{k-2}{k}u_0 e^{-u_0}\right) + \log\frac{4\pi}{k} \tag{6.14}$$

is nonnegative. Since the expression under the logarithm in (6.14) is decreasing with respect to u_0 and d ranges from 0 to 1, it attains its largest value for the minimal value of u_0, i.e. $3/\sqrt{8}$. So, the expression under the logarithm in (6.14) can be bounded by 0.97, which reduces this expression to $k/33 + \log(4\pi/k)$ and which is obviously greater than zero for all k.

For $3 \leqslant k \leqslant 7$ and $0 < d < 1$ we have $u_0 \in (0,1)$. So, $\psi_k(u)$ is increasing on $(1,\infty)$. Hence, to prove (6.13) in $(1,\infty)$ we omit the second term in (6.13) and it remains to check that

$$\psi_k(1) + \log\frac{4\pi}{k} = 2\sqrt{k} + 2 - k\log\left(1 + \frac{2e}{k} + \frac{2}{\sqrt{k}}\right) + \log\frac{4\pi}{k} \geqslant 0,$$

which we do directly for each value of k under consideration.

For $u \in (u_0,1)$ we shall consider $k = 5,6,7$ and $k = 3,4$, separately.

For $k = 5,6$ and 7 we see by a routine calculation that the expression $\theta = e^{-u} + \sqrt{k}u^{1+d}e^{-u}$ is either increasing on $(u_0,1)$ or increasing on some interval $\big(u_0, u_1(d,k)\big)$ and decreasing on $\big(u_1(d,k),1\big)$. Hence it attains the minimal value at $u=1$ or at $u=u_0$. As $u_0 \in (0,1)$, a simple calculation shows that θ is at least $5e^{-1}/2$. This means that the second term in (6.13) can be bounded from below by $2\log(1+5e^{-1}/2) > 1.3$. Again by the monotonicity of ψ_k on $(u_0,1)$ we infer that it is enough to check that

$$\begin{aligned} &-k\log\left(\frac{2}{k} + e^{-u_0} + \frac{k-2}{k}u_0e^{-u_0}\right) + 1.3 + \log\frac{4\pi}{k} \\ &\qquad\qquad \geqslant -k\log\left(1+\frac{2}{k}\right) + 1.3 + \log\frac{4\pi}{k} \geqslant 0. \end{aligned} \tag{6.15}$$

A straightforward calculation shows (6.15) for $k = 5,6$ and 7.

For $k = 3$ and 4 the second term on the left-hand side of (6.13) in $(u_0,1)$ is larger than $2\log(1+e^{-1}) > 0.6$. Arguing as above we see that it is enough to check that $-k\log(1+2/k) + 0.6 + \log(4\pi/k) \geqslant 0$, which obviously holds for $k = 3$ and 4. This concludes the proof. ■

Similarly to the reasoning for the upper bound and due to (3.2) in order to show that $u(\alpha,k) \geqslant u^*$ for some u^* it is enough to prove that

$$\frac{1-e^{-2}}{2}\,\frac{u^*}{u^*-k+2\sqrt{k}}\,\mathcal{E}_k(u^*) \geqslant \alpha. \tag{6.16}$$

For the future use we put $\kappa = -2\log\big((1-e^{-2})/2\big) \approx 1.6771$.

P r o o f o f P r o p o s i t i o n 5.1. First consider $k \geqslant 6$. Inserting $u^* = k + 2t - 5/2$ into (6.16) and taking logarithms of both sides we see that it is enough to prove

$$k \log\left(1 + \frac{2t - 5/2}{k}\right) - 2\log\left(1 + \frac{t - 5/4}{\sqrt{k}}\right) \geqslant \kappa + \log 4 - 5/2 \tag{6.17}$$

for $k \geqslant 6$ and $t \geqslant -\log 0.17$. Since for each $k \geqslant 6$ the left-hand side of (6.17) is increasing with respect to t and for $t = -\log 0.17$ is increasing with respect to k, we only check that

$$6\log\left(1 - \frac{2\log 0.17 + 5/2}{6}\right) - 2\log\left(1 - \frac{\log 0.17 + 5/4}{\sqrt{6}}\right) \geqslant \kappa + \log 4 - 5/2$$

by a direct calculation.

For $k = 2$ the inequality (5.1) is trivially satisfied. For $k = 3,\ 4$, and 5 we use (3.1) rather than (3.2) and instead of (6.16) and (6.17) we need to check that

$$(k-2)\log\left(1 + \frac{2t - 5/2}{k}\right) \geqslant \log 4k - 5/2 \tag{6.18}$$

for $t \geqslant -\log 0.17$. Since again the left-hand side of (6.18) is increasing with respect to t, we insert $t = -\log 0.17$ and verify (6.18) by a straightforward calculation. This completes the proof. ■

P r o o f o f T h e o r e m 5.1. Let us put $u^* = k + 2t + \sqrt{k}\log t$. Inserting u^* into (6.16), taking logarithms of both sides, setting $v = t/k$ and rearranging we reduce (6.16) to

$$k\log\left(1 + 2v + \frac{\log kv}{\sqrt{k}}\right) - \sqrt{k}\log kv - 2\log(2 + 2\sqrt{k}v + \log kv) - \kappa \geqslant 0. \tag{6.19}$$

A routine (although laborious) calculation shows that at least for every $k \geqslant 12$ the left-hand side of (6.19) is increasing with respect to v in the interval $[1/8, \infty)$. So, for the first case, it is enough to check that

$$k\log\left(\frac{5}{4} + \frac{\log(k/8)}{\sqrt{k}}\right) - \sqrt{k}\log\frac{k}{8} - 2\log\left(2 + \frac{\sqrt{k}}{4} + \log\frac{k}{8}\right) - \kappa \geqslant 0. \tag{6.20}$$

Using the facts that the function $\zeta(y) = (\log y)/y$ is decreasing on the interval (e, ∞), that the expression $(\log y)/\sqrt{y}$ attains maximal value $2e^{-1}$ and $k \geqslant 32$, we can write

$$2\log\left(2 + \frac{\sqrt{k}}{4} + \log\frac{k}{8}\right) \leqslant 2\zeta(2 + \sqrt{2} + \log 4)\left[\frac{1}{4} + \frac{2}{\sqrt{k}} + \frac{\log(k/8)}{\sqrt{k}}\right]\sqrt{k}$$
$$\leqslant \zeta(2 + \sqrt{2} + \log 4)\frac{1 + \sqrt{2} + \sqrt{8}e^{-1}}{2}\sqrt{k} \leqslant 0.565\sqrt{k}.$$

Hence and from the inequality $\log(1+y) \geqslant y - y^2/2$ the relation (6.20) can be reduced to

$$k\log\frac{5}{4} - \frac{1}{5}\sqrt{k}\log\frac{k}{8} - \frac{8}{25}\log^2\frac{k}{8} - 0.565\sqrt{k} - \kappa$$
$$\geqslant \sqrt{k}\Big(\sqrt{k}\log\frac{5}{4} - \Big[\frac{1}{5} + \frac{\sqrt{32}}{25e}\Big]\log\frac{k}{8} - 0.565\Big) - \kappa$$
$$\geqslant 32\log\frac{5}{4} - \Big[\frac{\sqrt{32}}{5} + \frac{32}{25e}\Big]\log 4 - 0.565\sqrt{32} - \kappa > 0,$$

which clearly holds true. This completes the proof of the first case. The second case, i.e. $k \geqslant 18$ and $v \geqslant 1/3$, can be proved in a similar way. ∎

Proof of Theorem 5.2. Proceeding in a similar way to that in (6.17) but for $u^* = k + 2t + \sqrt{kt}/4$ we reduce (5.3) to the following inequality:

(6.21)
$$\xi_k(v) = k\log\Big(1 + 2v + \frac{\sqrt{v}}{4}\Big) - \frac{k\sqrt{v}}{4} - \log k - 2\log\Big(\frac{2}{\sqrt{k}} + 2v + \frac{\sqrt{v}}{4}\Big) \geqslant \kappa$$

for $k \geqslant 17$ and $v \in [(\log 17)/k, 560]$. A routine calculation shows that for every $k \geqslant 17$, $\xi_k(v)$ is increasing on some interval $\big(v_1(k), v_2(k)\big)$, $0 < v_1(k) < 9 < v_2(k)$, and decreasing on the complementary intervals in $(0,\infty)$. Moreover, for $v = 0$ as well as for $v \to \infty$ the relation (6.21) is not satisfied. So, if for some $v_1^*(k) < v_2^*(k)$ we shall have $\xi_k\big(v_i^*(k)\big) > \kappa,\ i = 1, 2$, then (6.21) holds also in $[v_1^*(k), v_2^*(k)]$.

Now, we check that for all $k \geqslant 17$ we can take $v_2^*(k) = 560$. Indeed, (6.21) for $v = 560$ has the form

$$k[\log(1121 + \sqrt{35}) - \sqrt{35}] - 2\log(2 + [1120 + \sqrt{35}]\sqrt{k}) \geqslant \kappa.$$

Since the left-hand side of the above inequality is increasing with respect to k and for $k = 17$ the inequality holds true, our claim is proved.

Finally, we show that we can take $v_1^*(k) = (\log 17)/k$. In other words, we need to prove that for $k \geqslant 17$

$$\text{(6.22)}\quad \log 17\frac{k}{\log 17}\Big[\log\Big(1 + 2\frac{\log 17}{k} + \frac{1}{4}\sqrt{\frac{\log 17}{k}}\Big) - \frac{1}{4}\sqrt{\frac{\log 17}{k}}\Big]$$
$$- \log 17 - 2\log\Big(\frac{2}{\sqrt{k}} + 2\frac{\log 17}{k} + \frac{1}{4}\sqrt{\frac{\log 17}{k}}\Big) \geqslant \kappa.$$

Clearly, for $k \geqslant 17$ we have $(\log 17)/k \leqslant (\log 17)/17 \leqslant 0.1667$. Consequently, the last term on the left-hand side of (6.22) is greater than 0.165. Moreover, the first term is increasing with respect to k. Indeed, this follows since the function $[\log(1+2v+\sqrt{v}/4)-\sqrt{v}/4]/v$ is decreasing on $(0,1)$. So, inserting $k=17$ to the first term of (6.22) we see that the left-hand side is greater than 1.741, which concludes the proof. ■

7. APPENDIX

In this section, as a byproduct of our study we propose – using the results of previous sections – simple, explicit approximation formulae, which are easy to calculate for the normal and chi-square quantiles.

Normal quantiles. Putting $w=\sqrt{2t}=\sqrt{2\log(1/\alpha)}$ one can give an approximation $\hat{z}_\alpha$ for the normal quantiles z_α by fitting constants A and B in the formula (cf. (2.2))

$$\hat{z}_\alpha = w - \frac{A\log w + B}{w}. \tag{A.1}$$

We propose to take $A=0.866$ and $B=1.192$. With these constants (A.1) gives for $\alpha\in[0.000001, 0.1]$ a quite precise approximation as is shown in Table 1. For larger or smaller values of α this approximation also works although is less accurate.

TABLE 1. Approximation of the normal quantiles

α	z_α	$\hat{z}_\alpha$
0.1	1.2816	1.2824
0.05	1.6449	1.6441
0.01	2.3263	2.3253
0.005	2.5758	2.5751
0.001	3.0902	3.0903
0.0001	3.7190	3.7203
0.00001	4.2649	4.2671
0.000001	4.7534	4.7564

We have fitted the constants A and B to obtain a minimal value of the maximal relative error of approximation in the range [0.000001, 0.1], which turned out to be not greater than 0.065%. Therefore, our constants differ somewhat from those obtained by the least squares method.

Chi-square quantiles. Analysing the form of our lower and upper bounds from Sections 4 and 5 it is seen that besides the common terms k and $2t$ the following two different main terms appear: $\sqrt{kt}$ and $\sqrt{k}\log t$. So, we propose an approximation formula containing both such terms, i.e. taking the form

$$\hat{u}(\alpha,k) = k + 2t + A\sqrt{kt} + B\sqrt{k}\log t. \tag{A.2}$$

Observe that in the well-known Fisher approximation (see e.g. Zar [10]) the main terms are $k+t+2\sqrt{kt}$ while the Wilson–Hilferty formula (Zar [10]) has the main terms $k + \frac{4}{3}t + 2\sqrt{kt}$. In both formulae the term $\sqrt{k}\log t$ is absent and the term $2t$ is included only in part. This is the reason that those formulae give worse approximations for large t, i.e. for very small α.

To get more accurate approximation at least for typical range of α in the interval $[0.0001, 0.2]$ and typical values of k from 3 to 100 we complete (A.2) with some lower order terms. In effect we seek a formula of the form

$$\hat{u}(\alpha, k) = k + 2t + A\sqrt{kt} + B\sqrt{k}\log t + C\sqrt{k} + D\sqrt{t} + E\log t + F.$$

By fitting constants we propose finally the approximation formula of the form

$$\text{(A.3)} \quad \hat{u}(\alpha, k) = k + 2t + 1.62\sqrt{kt} + 0.63012\sqrt{k}\log t - 1.12032\sqrt{k} - 2.48\sqrt{t} - 0.65381\log t - 0.22872.$$

Numerical calculations show that the accuracy of (A.3) is comparable to that of the Wilson–Hilferty formula in the considered range of α and k. Moreover, (A.3) is better for small α while the competing formula is better for large k. If we restrict ourselves to $k \geqslant 6$, the relative error of approximation is not greater than 0.3% and in most cases does not exceed 0.1%.

REFERENCES

[1] Y. Baraud, S. Huet and B. Laurent, *Adaptive tests of linear hypotheses by model selection*, Ann. Statist. 31 (2003), pp. 225–251.

[2] C. W. Brain and J. Mi, *On some properties of the quantiles of the chi-square distribution and their applications to interval estimation*, Comm. Statist. Theory Methods 30 (2001), pp. 1851–1867.

[3] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh 1925.

[4] T. Inglot and T. Ledwina, *Asymptotic optimality of a new adaptive test in regression model*, Ann. Inst. H. Poincaré 42 (2006), pp. 579–590.

[5] C. Ittrich, D. Krause and W.-D. Richter, *Probabilities and large quantiles of noncentral chi-square distributions*, Statistics 34 (2000), pp. 53–101.

[6] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, Wiley, 1994.

[7] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 2, Wiley, 1995.

[8] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist. 28 (2000), pp. 1302–1338.

[9] E. B. Wilson and M. M. Hilferty, *The distribution of chi-square*, Proc. Nat. Acad. Sci. USA 17 (1931), pp. 684–688.

[10] J. H. Zar, *Approximations for the percentage points of the chi-square distribution*, Appl. Statist. 27 (1978), pp. 280–290.

Institute of Mathematics and Informatics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Tadeusz.Inglot@pwr.wroc.pl

Received on 23.11.2009;
revised version on 6.7.2010