

# ON INEQUALITIES THAT CONTAIN BOTH SAMPLE SIZES AND STUDENT T-QUANTILES DEPENDING ON THE SIZES: A BEAUTY OF SELECTED DIOPHANTINE PROBLEMS IN STATISTICS

Lubomír Štěpánek – Filip Habarta – Ivana Malá – Luboš Marek

## Abstract

There are problems in statistics that require solving an inequality that contains both sample size as an unknown variable and Student  $t$ -quantile that is dependent on the sample size. The dependency between the sample size and Student  $t$ -quantile is a function that cannot be expressed in any implicit form, and, thus, solving the given inequality, i. e. finding a minimal existing sample size that the inequality holds for, is tricky. Since the only acceptable solutions are natural numbers, those inequalities typical for statistical problems are of Diophantine kind. In this study, we define a general form of inequality for selected problems, discuss general solutions to such a form of inequalities, and suggest some simple numerical approaches to solving the inequalities, particularly using numerical algorithms and tabulation. Besides others, we also compare the estimates using standard normal quantiles and Student  $t$ -quantiles, which are usually more appropriate, to show differences between them.

**Keywords:** Student  $t$ -quantiles, standard normal quantiles, numerically-only solvable inequalities, Diophantine inequalities, tabulation of sample size-adjusted quantiles

**JEL code:** C12, C14, C18

## Introduction

In a wide variety of numerical problems in statistics, there are also those, usually called Diophantine problems, that natural numbers or integers are their only acceptable solutions (Brüderer & Dietmann, 2012).

As a typical example, one of the statistical problems' classes, when a natural number's solution is expected, is a minimum required sample size that ensures a confidence interval width of a selected estimate, calculated using the sample's values, is less than or equal to

a given constant. Supposing all hyperparameters of such a problem, besides the constant maximal (half)width of the estimate's confidence interval, are given, i. e. particularly a confidence level, then a goal is to find the minimal natural number so that the width of the confidence interval for the given estimate is not greater than the constant. Usually, under common (asymptotic) assumptions typical for the frequentist statistics such as the limit theorems (Skorokhod, 1956), we may assume the prior distribution of the point estimate follows normal or Student  $t$ -distribution (Lange et al, 1989). Thus, there are standard normal quantiles or Student  $t$ -quantiles usually presented in inequalities for the estimates' confidence intervals. However, whereas the standard normal quantiles are constant for a given confidence level, the Student  $t$ -quantiles depend not only on the confidence level but also on the sample size (Martin, 2012). That being said, there are both the sample size and the Student  $t$ -quantile (when used instead of the standard normal quantile) in one given inequality considered to solve the problem. Since the function describing the dependence of the Student  $t$ -quantile on the sample size cannot be expressed in an implicit form (Ng, 1988), solutions of such inequalities can be computed numerically only (Ohara & Sasaki, 2001). That may also increase the solutions' difficulty of the mentioned class of Diophantine inequalities in statistics; however, it opens room for non-routine approaches of the solutions (Baker, 1986).

Similarly, another class of problems that might be solved using an inequality that contains both the sample size and the Student  $t$ -quantile, which depends on the sample size, is an estimate of minimal sample size needed to reject the null hypothesis using an inference test's statistics. The null hypothesis rejection follows if and only if the test statistic is greater than the appropriate quantile with a given confidence level, either the standard normal one or the Student  $t$ -one. That condition may be reformulated analogously as for the first example with confidence intervals – the null hypothesis is rejected if and only if the sample size is large enough that a modified test statistic is greater than a derived constant under a given confidence level (Davenport & Roth, 1955). Thus, this is, in fact, an example of Diophantine inequality for another selected statistical problem.

This study addresses the mentioned issues with minimal sample size estimation to either keep the confidence interval's width lower than a given constant or ensure a slightly modified inference test statistic is greater than a given constant. The tricky part is that the Diophantine inequalities that solve the tasks contain both the sample size and Student  $t$ -quantile, a non-implicit function of the size. Therefore, we propose general refining of the tasks together with a general formulation of an appropriate Diophantine inequality. By suggesting a numerical approach and tabulation of specially defined terms, general and exact

solutions are provided. We also discuss how the usage of standard normal quantiles instead of the Student  $t$ -quantiles may lead to inaccurate estimation of the sample size.

## 1 Standard normal quantiles and Student $t$ -quantiles

Assuming a random variable  $Z$  follows the standard normal distribution,  $Z \sim N(0, 1^2)$ , then the  $(1 - \alpha/2)$ -th standard normal quantile is  $z_{1-\alpha/2}$ , so that

$$P(Z \leq z_{1-\alpha/2}) = F_Z(z_{1-\alpha/2}) = 1 - \alpha/2, \quad (1)$$

where the term  $P(\cdot)$  stands for a probability,  $F_Z$  is the cumulative distribution function of the variable  $Z$  and  $\alpha \in (0, 1)$  is the confidence level. Thus, the  $(1 - \alpha/2)$ -th standard normal quantile  $z_{1-\alpha/2}$  is then equal to

$$z_{1-\alpha/2} = F_Z^{-1}(F_Z(z_{1-\alpha/2})) = F_Z^{-1}(1 - \alpha/2). \quad (2)$$

The cumulative distribution function  $F_Z$  is necessary to evaluate the quantile  $z_{1-\alpha/2}$  and is equal to Newton definite integral of the probability mass function  $f_Z$  from  $-\infty$  to  $z_{1-\alpha/2}$ . So using the formula (1) we get

$$F_Z(z_{1-\alpha/2}) = \int_{-\infty}^{z_{1-\alpha/2}} f_Z(x) dx = \int_{-\infty}^{z_{1-\alpha/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \quad (3)$$

however, the definite integral from formula (3) can be simplified and evaluated not analytically, only numerically (Lukacs & King, 1954). Consequently, the  $(1 - \alpha/2)$ -th standard normal quantile  $z_{1-\alpha/2}$  from the formula (2) cannot be expressed implicitly, since the  $F_Z^{-1}$  function cannot be derived analytically and is usually tabulated or calculated numerically using statistical software.

Similar to the standard normal distribution case, assuming a random variable  $T$  follows Student  $t$ -distribution,  $T \sim T(\vartheta)$ , then the  $(1 - \alpha/2)$ -th Student quantile is  $t_{1-\alpha/2}(\vartheta)$  for  $\vartheta \in \mathbb{N}$  degrees of freedom. Keeping other mathematical notation the same as above, the cumulative distribution function  $F_T$  is equal to Newton definite integral of the probability mass function  $f_T$  from  $-\infty$  to  $t_{1-\alpha/2}(\vartheta)$ . So similarly to the formula (3) we get

$$F_T(t_{1-\alpha/2}(\vartheta)) = \int_{-\infty}^{t_{1-\alpha/2}(\vartheta)} f_T(x) dx = \int_{-\infty}^{t_{1-\alpha/2}(\vartheta)} \frac{\Gamma(\frac{\vartheta+1}{2})}{\sqrt{\vartheta\pi}\Gamma(\frac{\vartheta}{2})} \left(1 + \frac{x^2}{\vartheta}\right)^{-\frac{\vartheta+1}{2}} dx, \quad (4)$$

where  $\Gamma$  is the gamma function,  $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$  for any  $s \in \mathbb{R}$ . Although, the definite integral from formula (4) can be again evaluated only numerically (Chen et al, 2019). Therefore, the  $(1 - \alpha/2)$ -th Student quantile  $t_{1-\alpha/2}(\vartheta)$  for  $\vartheta \in \mathbb{N}$  degrees of freedom cannot

be expressed in the implicit form (the  $F_T^{-1}$  function cannot be derived analytically), and, thus, similarly as for the standard normal quantile, is tabulated or calculated numerically using appropriate software. Thus, Diophantine inequalities using standard normal or Student  $t$ -quantiles can usually be solved only numerically.

## 2 Selected classes of statistical problems solved using Diophantine inequalities

In the following subsections, we briefly recapitulate selected sorts of problems which solutions are available using Diophantine inequalities of our interest.

### 2.1 Searching for minimum sample size that ensures confidence interval's width of an estimated parameter is small enough

In this subsection, we assume a class of a confidence interval estimate for a sample parameter such that its halfwidth  $\Delta$  is proportional to a term  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}$ , where  $n$  is the sample size, and  $\kappa_{1-\alpha/2}$  is either  $(1 - \alpha/2)$  standard normal quantile,  $z_{1-\alpha/2}$ , or Student  $t$ -quantile,  $t_{1-\alpha/2}(\vartheta)$ , respectively, for a given  $\alpha \in (0, 1)$ . So, if the term  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}$  increases, the confidence interval's (half)width  $\Delta$  increases, too, and vice versa. The confidence interval halfwidth  $\Delta$  might be a function of other terms  $\delta_1, \delta_2, \delta_3, \dots$ , independent on the term  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}$ . Thus, in other words,

$$\Delta = f\left(\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}, \delta_1, \delta_2, \delta_3, \dots\right) \quad \text{and} \quad \Delta \propto \frac{\kappa_{1-\alpha/2}}{\sqrt{n}}. \quad (5)$$

Furthermore, in case there is  $\kappa_{1-\alpha/2} \equiv t_{1-\alpha/2}(\vartheta)$  in formula (6), the degrees of freedom  $\vartheta$  also depend on the sample size  $n$ , so  $\vartheta \propto n$  and, therefore,  $t_{1-\alpha/2}(\vartheta) \propto n$ . Eventually, both the numerator and denominator of the fraction  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}}$  depends on the sample size  $n$ .

The searching for the minimum sample size  $n$  that guarantees the halfwidth of the parameter's confidence interval would not be larger than  $\Delta$  may be defined as searching for minimal  $n$  that

$$f\left(\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}, \delta_1, \delta_2, \delta_3, \dots\right) \leq \Delta. \quad (6)$$

In case there is  $\kappa_{1-\alpha/2} \equiv z_{1-\alpha/2}$  in inequality (6), since the terms  $\delta_1, \delta_2, \delta_3, \dots$  are independent on the term  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}}$ , the denominator  $\sqrt{n}$  can be isolated and the (Diophantine) inequality (6) may be rewritten as  $f\left(\frac{z_{1-\alpha/2}}{\Delta}, \delta_1, \delta_2, \delta_3, \dots\right) \leq \sqrt{n}$ . That enables to easily

estimate the minimum value of  $n$ . However, in case there is  $\kappa_{1-\alpha/2} \equiv t_{1-\alpha/2}(\vartheta)$  in inequality (6), the tricky part is that not only the denominator of  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}}$ , but also its numerator depends on  $n$ . So, simple isolation of  $n$  could not work to get an exact value of  $n$ .

As an example of this confidence intervals' class, a halfwidth of a confidence interval for a population mean follows a form  $\Delta = \frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \cdot s$ , where  $s$  is a sample standard deviation. Similarly, a halfwidth of a confidence interval for a population proportion follows a form  $\Delta = \frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \cdot \sqrt{p(1-p)}$ , where  $p$  is a sample proportion. Thus, they do not violate the assumptions in (5) and (6). The task for the first example is to find minimal  $n$  so that  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \cdot s \leq \Delta$ , and for the second is to find minimal  $n$  so that  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \cdot \sqrt{p(1-p)} \leq \Delta$ . While the solution for the standard normal quantiles is easy following the (6) and rewriting the inequalities as  $\frac{z_{1-\alpha/2}}{\Delta} \cdot s \leq \sqrt{n}$  or  $\frac{z_{1-\alpha/2}}{\Delta} \cdot \sqrt{p(1-p)} \leq \sqrt{n}$ , in case of the Student  $t$ -quantiles, such isolation is not helpful (and possible), since Student  $t$ -quantile  $t_{1-\alpha/2}(\vartheta)$  depends on sample size  $n$  as  $\vartheta = n - 1$ , so the halfwidth  $\Delta$  is  $\Delta = \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \cdot s = \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}} \cdot s$  or  $\Delta = \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \cdot \sqrt{p(1-p)} = \frac{t_{1-\alpha/2}(n-1)}{\sqrt{n}} \cdot \sqrt{p(1-p)}$ , respectively.

## 2.2 Searching for minimum sample size that ensures inference test's statistic is greater than or equal to appropriate quantile

We take into account for all inference tests with a test's statistic  $s$  as a function  $s = f(\sqrt{n}, \delta_1, \delta_2, \delta_3, \dots)$ , where  $n$  is a sample size and  $\delta_1, \delta_2, \delta_3, \dots$  are terms independent on the sample size  $n$ , and the statistic  $s$  is proportional to  $\sqrt{n}$ , i. e.  $s \propto \sqrt{n}$ . Assuming  $\alpha \in (0, 1)$  is a confidence level, then, usually, if the test' statistic  $s$  is greater than or equal to an appropriate  $(1 - \alpha/2)$ -th quantile, either the standard normal one,  $z_{1-\alpha/2}$ , or the Student  $t$ -quantile,  $t_{1-\alpha/2}(\vartheta)$ , the null hypothesis is rejected. Considering all arguments of the function  $s = f(\sqrt{n}, \delta_1, \delta_2, \delta_3, \dots)$  remain roughly constant with increasing sample size  $n$ , a typical task is to estimate minimal sample size  $n$  that would result into the null hypothesis rejection. Thus, the task is to find minimal  $n$  so that

$$s = f(\sqrt{n}, \delta_1, \delta_2, \delta_3, \dots) \geq \kappa_{1-\alpha/2}, \quad (7)$$

which, since  $\sqrt{n}$  is not dependent on  $\delta_1, \delta_2, \delta_3, \dots$ , can be rewritten as

$$\frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \leq f(\delta_1, \delta_2, \delta_3, \dots) = s. \quad (8)$$

In case of  $\kappa_{1-\alpha/2} \equiv z_{1-\alpha/2}$  in Diophantine inequality (8), the solution is made by isolation of  $\sqrt{n}$  as following,  $\frac{z_{1-\alpha/2}}{s} \leq \sqrt{n}$ . However, such an isolation of  $\sqrt{n}$  could not work for exact solution of the inequality (8) if  $\kappa_{1-\alpha/2} \equiv t_{1-\alpha/2}(\vartheta)$ , since both the numerator and denominator of the term  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}}$  depends on  $n$ .

As an example of an inference test of this class, the classical one-sample  $t$ -test uses test statistic following a form of  $s = f(\sqrt{n}, \delta_1, \delta_2, \delta_3, \dots)$ , where  $n$  is a sample size and  $\delta_1, \delta_2, \delta_3, \dots$  are term independent of the sample size. The null hypothesis, claiming there is no statistical difference between the population mean and a constant, is rejected at the confidence level  $\alpha \in (0, 1)$ , if and only if the test statistic is greater than an appropriate quantile,  $\kappa_{1-\alpha/2}$ ; thus, when  $s = f(\sqrt{n}, \delta_1, \delta_2, \delta_3, \dots) \geq \kappa_{1-\alpha/2}$ . The task is then to find minimal  $n$  so that  $\frac{\kappa_{1-\alpha/2}}{\sqrt{n}} \leq f(\delta_1, \delta_2, \delta_3, \dots) = s$  for  $\kappa_{1-\alpha/2} \in \{z_{1-\alpha/2}, t_{1-\alpha/2}(\vartheta)\}$ , where  $\vartheta = n - 2$  stands for degrees of freedom and  $\alpha \in (0, 1)$  is the confidence level.

### 3 Proposed approaches to solutions of the defined Diophantine inequalities

When we assume usage of a standard normal quantile, so there is  $\kappa_{1-\alpha/2} \equiv z_{1-\alpha/2}$  is formulas (6) and (7), the formulas could be rewritten as  $f\left(\frac{z_{1-\alpha/2}}{\Delta}, \delta_1, \delta_2, \delta_3, \dots\right) \leq \sqrt{n}$  and,  $\frac{z_{1-\alpha/2}}{f(\delta_1, \delta_2, \delta_3, \dots)} \leq \sqrt{n}$ , respectively. Since the terms  $z_{1-\alpha/2}, \delta_1, \delta_2, \delta_3, \dots$  are all independent on  $n$ , we can exactly find such  $n \in \mathbb{N}$  that solves the inequalities.

However, if we assume Student  $t$ -quantiles,  $\kappa_{1-\alpha/2} \equiv t_{1-\alpha/2}(\vartheta)$ , there is no possibility to isolate  $\sqrt{n}$  in formulas (6) and (7), since the term  $t_{1-\alpha/2}(\vartheta)$  also depends on  $n$  due to degrees of freedom,  $\vartheta$ , which are a linear function of  $n$ ,  $\vartheta = f(n)$ , typically as  $\vartheta = n - 1$  or  $\vartheta = n - 2$ . As a consequence, an analytical solution, as in the previous paragraph, is not possible here. Although, thanks to the assumed proportionality  $\Delta \propto \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}}$ , we can reformulate the formula (6) as  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \leq \frac{\Delta}{f(\delta_1, \delta_2, \delta_3, \dots)} = \text{const}$  and formula (7) as  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \leq \frac{s}{f(\delta_1, \delta_2, \delta_3, \dots)} = \text{const}$ . So, we get a Diophantine inequality as follows,

$$\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \leq r = \text{const.} > 0, \quad (9)$$

from which we get  $\frac{t_{1-\alpha/2}(\vartheta)}{r} \leq \sqrt{n}$ .

### 3.1 Usage of standard normal quantiles and Student $t$ -quantiles

Usually, we use standard normal quantiles in statistical problem enabling to assume asymptotic properties of applied statistics are met, e. g. the sample size is sufficient and estimates used to evaluate the statistics are known in advance from literature or large populations. But, in fact, this is hardly satisfied using real-world data. Thus, Student  $t$ -quantile should be preferred before the standard normal ones, even if they may challenge the above-mentioned problems.

As for illustration, assuming  $\alpha = 0.05$  and  $\vartheta = f(n) = n - 1$ , we got different estimates of minimal  $n$  satisfying the inequality (9), is shown in Table 1. Thus, using the standard normal quantiles tends to underestimate the minimal  $n$ .

**Table 1: Table of minimal  $n$  estimates using both standard normal quantiles and Student  $t$ -quantiles for  $\alpha = 0.05$  and  $\vartheta = f(n) = n - 1$**

$r$	0.70	0.50	0.30	0.10	0.05
estimated minimal $n$ using $z_{0.975}$	8	16	43	385	1537
estimated minimal $n$ using $t_{0.975}(n - 1)$	11	18	46	387	1540

### 3.2 A numerical solution of the defined Diophantine inequalities

As the first estimate of  $n \in \mathbb{N}$  in (9), we can use that for each  $\vartheta$ , it is  $z_{1-\alpha/2} \leq t_{1-\alpha/2}(\vartheta)$ ; thus, using (9), it is also  $\frac{z_{1-\alpha/2}}{r} \leq \frac{t_{1-\alpha/2}(\vartheta)}{r} \leq \sqrt{n}$  and  $n \geq \left(\frac{z_{1-\alpha/2}}{r}\right)^2$ . Finally, we get the first estimate using the ceiling function<sup>1</sup> as  $n = \left\lceil \frac{z_{1-\alpha/2}}{r} \right\rceil^2$  which obviously  $n = \left\lceil \frac{z_{1-\alpha/2}}{r} \right\rceil^2 \in \mathbb{N}$ .

Now let us prove a (Diophantine) solution  $n \in \mathbb{N}$  of inequality (9) always exists. Using the theorem of two bounding functions' limits, if  $\forall n \in \mathbb{N}$  it is  $\frac{0}{\sqrt{n}} \leq \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \leq \frac{\max_{\vartheta \in \mathbb{N}}\{t_{1-\alpha/2}(\vartheta)\}}{\sqrt{n}} < \infty$ , then  $\lim_{n \rightarrow \infty} \frac{0}{\sqrt{n}} = 0$  and  $\lim_{n \rightarrow \infty} \frac{\max_{\vartheta \in \mathbb{N}}\{t_{1-\alpha/2}(\vartheta)\}}{\sqrt{n}} = 0$ , it is also  $\lim_{n \rightarrow \infty} \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} = 0$ . Thus if  $\lim_{n \rightarrow \infty} \frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} = 0$ , then also  $\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} \leq r > 0$  and  $n \in \mathbb{R}$  must always exist. If  $n \in \mathbb{R}$  exists, then also  $[n] \in \mathbb{N}$  as Diophantine solution exists, holding  $[n] \geq n$ .

So, we suggest the following algorithm to numerically find the minimal  $n \in \mathbb{N}$  satisfying the Diophantine inequality (9), see Algorithm 1.

<sup>1</sup> The ceiling function maps  $x$  to the least integer greater than or equal to  $x$ , denoted  $[x]$ . So,  $[x] = \min_{z \in \mathbb{Z}}\{z \geq x\}$ .

**Algorithm 1: Numerical searching for the minimal  $n \in \mathbb{N}$  satisfying the Diophantine inequality (9)**

**Data:** a value of  $r > 0$ , confidence level  $\alpha \in (0, 1)$  and degrees of freedom  $\vartheta = f(n)$   
**Result:** numerical estimate of minimal  $n \in \mathbb{N}$  satisfying the Diophantine inequality (9)

```

1    $n = \left\lceil \frac{z_{1-\alpha/2}}{r} \right\rceil^2;$ 
2   while  $\frac{t_{1-\alpha/2}(f(n))}{\sqrt{n}} > r$  do
3        $n = n + 1;$ 
4    $n;$ 
    
```

**3.3 A solution using tabulation of the defined Diophantine inequalities**

Assuming the inequality (9) as equation, i. e.

$$\frac{t_{1-\alpha/2}(\vartheta)}{\sqrt{n}} = r = \text{const.} > 0, \tag{10}$$

the confidence level  $\alpha \in (0, 1)$ , and also that the degrees of freedom are a linear function of  $n$ ,  $\vartheta = f(n)$ , typically as  $\vartheta = n - 1$  or  $\vartheta = n - 2$ , we can simply tabulate the left-hand side of equation (10), i. e. the fractions of Student  $t$ -quantiles  $t_{1-\alpha/2}(f(n))$  and square root of  $n$ , i. e.  $\sqrt{n}$ . See the logic of the tabulation in Table 2. Using such a table, a numerical estimate of  $n \in \mathbb{N}$  can be found very precisely in an appropriate row of the first column the table, if  $r$  is found in the third column of the table. When  $t_{1-\alpha/2}(f(n))/\sqrt{n} < r < t_{1-\alpha/2}(f(n+1))/\sqrt{n+1}$ , then the minimal sample size is better taken as  $n + 1$ , not as  $n$ .

**Table 2: Tabulation of  $t_{1-\alpha/2}(f(n))/\sqrt{n}$  fractions for given degrees of freedom's function  $\vartheta = f(n)$  and confidence level  $\alpha \in (0, 1)$**

$n$	$\vartheta = f(n)$	$t_{1-\alpha/2}(f(n))/\sqrt{n}$
2	$f(2)$	$t_{1-\alpha/2}(f(2))/\sqrt{2}$
3	$f(3)$	$t_{1-\alpha/2}(f(3))/\sqrt{3}$
4	$f(4)$	$t_{1-\alpha/2}(f(4))/\sqrt{4}$
$\vdots$	$\vdots$	$\vdots$

**Conclusion remarks**

Estimation of minimum sample size, ensuring the parameter's confidence interval width will be low enough, or inference statistic would be large enough, should use Student  $t$ -quantiles rather than the standard normal ones not to underestimate the sample size. However, since



Student  $t$ -quantiles depend on the sample size via degrees of freedom, the estimation of minimum sample size is tricky and demands numerical solving of Diophantine inequality.

We defined the task and the inequality and suggested the numerical approach and tabulation of Student  $t$ -quantiles divided by sample sizes, both enabling finding the accurate minimal sample size estimate. While the numerical technique usually requires computational software and provides an exact solution using an initial estimate of minimal sample size via the standard normal quantile, the tabulation approach demands only paper-n-pencil and searching for the estimate in the derived table.

## Acknowledgment

This paper is supported by the grant OP VVV IGA/A, CZ.02.2.69/0.0/0.0/19\_073/0016936 (18/2021) by Internal Grant Agency of Prague University of Economics and Business.

## References

- Baker, R. (1986). *Diophantine inequalities*. Oxford Oxfordshire New York: Clarendon Press Oxford University Press
- Brüder, J., & Dietmann, R. (2012). Random Diophantine inequalities of additive type. *Advances in Mathematics*, 229(6), 3079–3095. <https://doi.org/10.1016/j.aim.2012.02.004>
- Davenport, H., & Roth, K. F. (1955). The solubility of certain Diophantine inequalities. *Mathematika*, 2(2), 81–96. <https://doi.org/10.1112/s0025579300000723>
- Chen, Z., Wang, B., & Gorban, A. N. (2019). Multivariate Gaussian and Student-t process regression for multi-output prediction. *Neural Computing and Applications*, 32(8), 3005–3028. <https://doi.org/10.1007/s00521-019-04687-8>
- Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust Statistical Modeling Using the Distribution. *Journal of the American Statistical Association*, 84(408), 881–896. <https://doi.org/10.1080/01621459.1989.10478852>
- Lukacs, E., & King, E. P. (1954). A Property of the Normal Distribution. *The Annals of Mathematical Statistics*, 25(2), 389–394. <https://doi.org/10.1214/aoms/1177728796>
- Martin, B. R. (2012). Sampling Distributions Associated with the Normal Distribution. In *Statistics for Physical Science* (pp. 105–122). Elsevier. <https://doi.org/10.1016/b978-0-12-387760-4.00006-8>

Ng, K.-F. (1988). An Inequality Implicit Function Theorem. *Journal of the Australian Mathematical Society. Series A. Pure Mathematics and Statistics*, 44(2), 146–150.  
<https://doi.org/10.1017/s1446788700029736>

Ohara, A., & Sasaki, Y. (2001). On solvability and numerical solutions of parameter-dependent differential matrix inequality. *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*. 40th Conference on Decision and Control.  
<https://doi.org/10.1109/cdc.2001.980417>

Skorokhod, A. V. (1956). Limit Theorems for Stochastic Processes. *Theory of Probability & Its Applications*, 1(3), 261–290. <https://doi.org/10.1137/1101022>

## Contact

Lubomír Štěpánek

Faculty of Informatics and Statistics, Prague University of Economics and Business

W. Churchill Sq. 1938/4, 130 67 Prague, Czech Republic

lubomir.stepanek@vse.cz

Filip Habarta

Faculty of Informatics and Statistics, University of Economics and Business

W. Churchill Sq. 1938/4, 130 67 Prague, Czech Republic

filip.habarta@vse.cz

Ivana Malá

Faculty of Informatics and Statistics, University of Economics and Business

W. Churchill Sq. 1938/4, 130 67 Prague, Czech Republic

malai@vse.cz

Luboš Marek

Faculty of Informatics and Statistics, University of Economics and Business

W. Churchill Sq. 1938/4, 130 67 Prague, Czech Republic

marek@vse.cz