COMPUTATION OF TYPE A UNCERTAINTY FROM THE NORMALLY DISTRIBUTED ORDER STATISTICS

Ivan JANIGA, Roman MIKLÓŠ, Rudolf PALENČÁR

Faculty of Mechanical Engineering, Slovak University of Technology
Námestie slobody 17, 812 31 Bratislava, Slovakia

janiga@sjf.stuba.sk, rmiklos@cvt.stuba.sk

Abstract

Metrological characteristics of a measurement instrument to be used in any statistical process control (more about SPC see in [6], [7], [8], [9], [10]) must be found out and regularly verified. Uncertainties (more about them see in [4], [5]) of the maximum deviation between a measured value on the measuring instrument and a value on the etalon, which are expressed as order statistics, are important. The type A uncertainty of the normally distributed (normality tests see in [11], [12]) order statistic can be expressed through the standard normal order statistic, the evaluation of which is given in the paper.

Key words: order statistic, statistical process control, type A uncertainty, standard normal distribution

1 Introduction

In statistical practice there is often needful to know the mean values and variances of certain random variables, so-called order statistics [2].

The definition of order statistics is given e.g. in [1], [2]. There are especially interesting mean values and dispersions of order statistics from the distribution $N(0,1)$. These values were never tabulated before in such detail (so far of the authors´ knowledge). There are only published values available for relatively small range of the random samples *n*.

Let us denote both $X_{[i]}$ $(i = 1,...,n)$ the *i*-th order statistic drawn from the population which has the continuous distribution function $F(x)$ and $U_{[i]}$ $(i = 1,...,n)$ the *i*-th order statistic drawn from the population which has the standardized normal distribution $N(0,1)$.

2 Theoretical analysis

The determination of mean values $E(U_{[i]})$ and dispersions $D(U_{[i]})$ requires the numerical computation of improper integrals in which both limits take on infinite values. The theorem below gives the sufficient condition for existence of mean values and variances of $X_{[i]}$, i.e. the sufficient condition for convergence of related improper integrals.

Theorem. *Let X be an arbitrary continuous random variable with density function f and distribution function F.*
If there exist the mean value $E(X)$ *and the variance* $D(X)$, *then* $\forall\, n \in N$ *there exist mean values* $E(X_{[i]})$ *and variances* $D(X_{[i]})$ *of all order statistics* $X_{[i]}$ $(i = 1,...,n)$.

Proof. The existence assumption and the definitions of mean value and variance of a continuous random variable X conclude automatically the convergence of related improper integrals, because

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx\,,\qquad D(X)=\int_{-\infty}^{\infty} x^2 f(x)dx-E^2(X) \tag{1}$$

The probability density function $f_{[i]}(x)=\mathrm{d}F_{[i]}(x)/\mathrm{d}x$ of statistic $X_{[i]}$ is given by the formula (see in [1]) :

$$f_{[i]}(x)=\frac{n!}{(i-1)!(n-i)!}\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i} f(x)\,,\qquad -\infty < \mathrm{x} < +\infty \tag{2}$$

Obviously, the property of distribution $0\le F(x)\le 1$ implies, that for $i=1,...,n$ and any $x\in \boldsymbol{R}$ the next is valid

$$\frac{(\mathrm{i}-1)!(\mathrm{n}-\mathrm{i})!}{n!} f_{[i]}(x)\le f(x) \tag{3}$$

and hence from properties of density $f(x)$ it follows the satisfaction of the next two conditions

$$\frac{(\mathrm{i}-1)!(\mathrm{n}-\mathrm{i})!}{n!}\left|xf_{[i]}(x)\right|\le\left|xf(x)\right| \tag{4}$$

$$\frac{(\mathrm{i}-1)!(\mathrm{n}-\mathrm{i})!}{n!} x^2 f_{[i]}(x)\le x^2 f(x)\,. \tag{5}$$

Consequently, convergence of integrals (1) and properties (4) and (5) conclude, according to the main criterion for convergence of improper integrals, the convergence of the following integrals

$$\int_{-\infty}^{\infty} xf_{[i]}(x)\mathrm{d}x\,,\qquad \int_{-\infty}^{\infty} x^2 f_{[i]}(x)\mathrm{d}x \tag{6}$$

Thus, it is appropriate to say, that for any $n\in \boldsymbol{N}$ and $i=1,...,n$ the existence of mean values $E(X_{[i]})$ and variances $D(X_{[i]})$ is established.

This theorem is for the general case and we therefore specialize it further by considering a special case in the following corollary, which establishes that the considered improper integrals are really convergent for $U_{[i]}$, thus their numerical treatment will deliver adequate results.

Corollary. *For any $n \in \mathbf{N}$ there exist mean values $E(U_{[i]})$ and variances $D(U_{[i]})$ of all order statistics $U_{[i]}$ ($i = 1, ..., n$).*

Proof. The assumption $U \sim N(0,1)$ implies immediately the existence of mean value $E(U) = 0$ and variance $D(U) = 1$, i.e. the sufficient condition for existence $E(U_{[i]})$ and $D(U_{[i]})$ is satisfied. Applying the theorem above gives the result.

3 Numerical computation

The computation of mean values $E(U_{[i]})$ and standard errors $\sigma(U_{[i]}) = \sqrt{D(U_{[i]})}$ was processed in the computer algebra system *Mathematica*, which supports various quadrature methods with practical arbitrary working precision. We used the adaptive *Gauss - Kronrod* method, which is a modification of the classical *Gauss* quadrature method.

In [3] are presented six tables with computed values rounded to four decimal places. In *table 1* and in *table 2* respectively are presented the computed values of $E(U_{[i]})$ and $\sigma(U_{[i]})$ for $n = 1, ..., 30$. The row number *n* gives the size of the random sample and the column is marked with *k*. In table headings is explained how to determine *i* in *n*-th row and *k*-th column.
In other tables are presented only values for *n*-th order statistic i.e. $E(U_{[n]})$ and $\sigma(U_{[n]})$ respectively. *Table 3* and *table 4* give values for *n=10,20,..,90,100,200,..,900,.., further* 1×10^4, 2×10^4, ..,9×10^4,.., 1×10^6, 2×10^6, ..,9×10^6. At last, *tables 5* and *6* contain values for *n=10,..,1000* with step 10. The determination of *n* in *i*-th row and *j*-th column is explained in heading of each table.

4 Order statistics and the type A uncertainty

The application of order statistics in a gauge testing process is given in detail in the article [2]. We only note that the *type A uncertainty* of the greatest observed deviation between the gauge measured value and the etalon's value can be expressed as

$$\sigma(X_{[n]}) = \sigma(U_{[n]}) \cdot \sigma \tag{7}$$

where σ is the standard deviation of an individual measurement *x*, which is usually unknown so it should be estimated from measurements. A biased estimator of the standard deviation σ is

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \; . \tag{8}$$

Then the *type A uncertainty* of the greatest deviation $x_{\max}$ is

$$u_{Ax_{\max}} = \sigma(U_{[n]}) \cdot \sigma \; . \tag{9}$$

At last, the determination of $u_{Ax_{\max}}$ according to above formulae (8) and (9) reduces to determination of $\sigma(U_{[n]})$ from the presented tables.

5 Conclusion

We described the approach to the numerical computation of mean values and standard deviations of order statistics, which can be of great importance in the statistical measurement process control.
Further we presented the concrete tables of mean values *E*(.) and standard deviations *σ*(.) of the following order statistics from the standard normal distribution $N(0,1)$:

- $U_{[i]}$ for $n = 1,...,30$; $(1 \le i \le n)$
- $U_{[n]}$ for very wide sample size *n* $(n \le 9\times10^6)$

Tables of these types and range were not published before.

Acknowledgement

This contribution was supported by the Grant No. 1/8094/01 as well as the Grant No. 1/7077/20 both from VEGA

Bibliography

[1] Janiga I.: Štatistické metódy pri overovaní meradiel. *Habilitačná práca*. SjF STU v Bratislave, 1998

[2] Janiga I., Palenčár R.: On Normal Distributed Deviation in Gauge Testing Process, MEASUREMENT '99 proceedings, Int. Conf. On Measurement Smolenice, 1999, 19-22

[3] Janiga I., Miklóš R.: Computation of some Order Statistics' Moments, Proceedings Int. Conf. Matematyka w naukach technicznych i przyrodniczych, Krynica, 4--7 czerwca 2000

[4] Palenčár R., Halaj M.: Metrologické zabezpečenie systémov riadenia kvality, STU Bratislava, 1998

[5] Chudý V., Palenčár R., Kureková E., Halaj M.: Meranie technických veličín, STU Bratislava, 1999

[6] Terek M., Hrnčiarová Ľ.: Possibilities of Economic Design of 3σ - Control Charts. Scientific Papers of the University of Pardubice 1998

[7] Terek M., Peller F.: Some Possibilities of the Analysis of the Costs of SPC by Control Charts. Scientific Papers of the University of Pardubice 1999

[8] Terek M., Hrnčiarová Ľ.: Štatistické riadenie kvality, Ekonom Bratislava 1999

[9] Hrnčiarová Ľ., Terek M.: Použitie nástrojov na zlepšenie kvality. Kvalita č.1/1998

[10] Hrnčiarová, Ľ., Terek, M.: Štatistické riadenie kvality: Indexy spôsobilosti a ich testovanie. Kvalita, Inovácia, Prosperita č. 1-2, 1999

[11] Garaj I.: W-kritérium a jeho spracovanie na počítači, 3, seminár výpočtová štatistika, Bratislava, 1990, str. 8-9

[12] Rublík F., Bognárová M.: Tables for a statistical control test, Applications of mathematics 37, No. 6, 1992