METROLOGY AND MEASUREMENT SYSTEMS
VOL. XV, NUMBER 4 (2008)

M. Catelani, A. Zanobini, L. Ciani

University of Florence, Department of Electronics and Telecommunications
Italy
e-mai: marcantonio.catelani@unifi.it

INTRODUCTION TO THE *T* AND *CHI-SQUARE* DISTRIBUTION FOR A MORE ACCURATE EVALUATION OF THE MEASURE OF THE WORD ERROR RATE IN ANALOG-TO-DIGITAL CONVERTERS

The word error rate (WER) in an Analog to Digital Converter (ADC) is the probability of receiving an erroneous code for an input, after correction is made for gain, offset, and nonlinearity errors, and a specified allowance is made for noise. Typical causes of word errors are metastability and timing jitter of comparators within the ADC [1].

New statistical techniques which can better integrate what is sustained in the IEEE standard and in [2] have been proposed. In particular, Student and chi-square distributions have been introduced for a more accurate measurement of the word error rate in the case of n successive observations.

1. RECALL OF THEORETICAL BASIS [3], [4]

It is well known that a generic normally distributed measure $M = N(m, u)$ with expected value m and standard uncertainty u, can be expressed in a reduced form using the $\frac{M-m}{u} = N(0,1)$ normally distributed, with expected value zero and unitary standard uncertainty. It is also known that in the presence of a number ν of normal random variables given in the reduced form $N_1(0,1), \cdots, N_\nu(0,1)$, mutually independent and independent from M, the sum of squares of such variables, that is $\chi_\nu^2 = \sum_{i=1}^{n} N_i^2(0,1)$ is distributed like a chi-square distribution with ν degrees of freedom. The ratio between the original measure expressed in the reduced form $\frac{M-m}{u} = N(0,1)$ and the positive square root of χ_ν^2 divided by degrees of freedom ν, can be expressed by the following variable:

[1] Received: November 3, 2008. Revised: November 21, 2008.

$$T_\nu = \frac{N(0,1)}{\sqrt{\sum_{i=1}^{\nu} N_i^2(0,1)\big/\nu}} \text{, with } E\left\{N_i(0,1), N_j(0,1)\right\} = 0 \tag{1}$$
$$i \neq j = 1, \cdots, \nu \text{(condition of independence).}$$

T_ν follows a Student distribution (or t distribution) with ν degrees of freedom. Its probability density function can be expressed by

$$f_\nu(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma[(\nu+1)/2]}{\Gamma[\nu/2]} \frac{1}{\left(1 + t^2\big/\nu\right)^{(\nu+1)/2}} \qquad -\infty < t < +\infty; \quad \nu = 1, 2, ..., \tag{2}$$

where $\Gamma(\alpha) = \int_0^{+\infty} e^{-x}\, x^{\alpha-1}\, dx$ represents the generic gamma function.

In the case of $\nu = 1$, Eq. (1) becomes the probability density function of the so-called Cauchy distribution for which neither the expected value nor the variance are defined.

It can be also demonstrated that:

$$\lim_{\nu\to\infty} f_\nu(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}; \quad \lim_{\nu\to\infty} Var\{T_\nu\} = \lim_{\nu\to\infty} \frac{\nu}{\nu-2} = 1\ , \tag{3}$$

that is, when the number of degrees of freedom tends to infinity, the Student variable tends to the normal reduced variable. The confidence level p and the corresponding uncertainty interval $[-t_p, +t_p]$, centred around zero for the Student variable with ν degrees of freedom, is determined as follows:

$$p = P\left\{-t_p \leq T_\nu \leq t_p\right\} = \int_{-t_p}^{t_p} f_\nu(t)\, dt. \tag{4}$$

The Student table of quantiles (see for instance GUM [5] – Table 2.2 p.66) offers different values of t_p for successive numbers of degrees of freedom corresponding to the various confidence levels p.

2. ESTIMATION OF THE WORD ERROR RATE FOR N SUCCESSIVE OBSERVATIONS

We can consider n independent successive observations $(o_1, \cdots, o_n)$ of the same measurand (trial samples), obtained by the same measurement process implemented in identical conditions of repeatability. As known we are, in this situation, in the presence of intrinsic random effects on the measurement process (*inherent variability of the measurement process*). If we hypothesize that each observation of the word error is a normally random variable with expected value m_o and standard uncertainty u_o, then $O_i = N(m_o, u_o) \quad \forall i = 1, ..., n$.

The arithmetic mean $\overline{O} = \sum_{i=1}^{n} \frac{o_i}{n} \approx N\left(m_o, \frac{u_0^2}{n}\right)$ is an estimator, that is a random variable with the same expected value m_o and variance reduced by a factor $(1/n)$.

We now introduce the normal arithmetic mean expressed in its reduced form, that is: $\frac{\overline{O} - m_o}{\frac{u_o}{\sqrt{n}}} = N(0, 1)$. An estimator $S\left(\overline{O}\right)$ of $\frac{u_o}{\sqrt{n}}$ exists, and it is given by the so-called experimental standard deviation of the mean, according to the following formula:

$$S\left(\overline{O}\right) = \sqrt{\sum_{i=1}^{n} \frac{\left(o_i - \overline{O}\right)^2}{[n(n-1)]}} = \frac{u_o}{\sqrt{n}} \sqrt{\frac{\chi^2_{n-1}}{n-1}}, \tag{5}$$

being:

$$\chi^2_{n-1} = \sum_{i=1}^{n} \frac{\left(o_i - \overline{O}\right)^2}{u_o^2} = \sum_{i=1}^{n} N_i^2(0, 1). \tag{6}$$

Eq. (6) denotes the well known chi-square with $(n - 1)$ degrees of freedom, with the $N_i(0, 1)$ mutually independent and independent from $\overline{O}$ and therefore from $N_o(0, 1)$.

Consequently, it can be verified that:

$$T_{n-1} = \frac{\overline{O} - m_o}{S\left(\overline{O}\right)} = \frac{N_o(0, 1)}{\sqrt{\chi^2_{n-1}/(n-1)}}, \tag{7}$$

is a Student variable with $(n - 1)$ degrees of freedom.

If now we consider the uncertainty interval introduced in (3), assuming $\nu = n\text{-}1$ and adopting (7), we can write:

$$p = P\left\{-t_p \le T_{n-1} \le t_p\right\} = P\left\{m_o - t_p S\left(\overline{O}\right) \le \overline{O} \le m_o + t_p S\left(\overline{O}\right)\right\}, \tag{8}$$

which represents an estimation of the confidence level shown in Eq. (4).

The arithmetic mean $\overline{O}$ can be interpreted as a final measure but in Eq. (8) any useful information about the uncertainty interval are not deduced being m_o generally unknown and the estimator $S\left(\overline{O}\right)$ of Eq. (5) a random variable.

An approximate solution can be obtained by introducing (see GUM [5] G.3.1) the Student variable:

$$T_{n-1} = \frac{M - \overline{o}}{s_{\overline{o}}}$$

where the observed word errors $[o_1, \cdots, o_n]$ are introduced. In this case $\overline{o}$ represents the estimate of m_o while $s_{\overline{o}} = \sqrt{\sum_{i=1}^{n} (o_i - \overline{o})^2 \Big/ [n\,(n-1)]}$ is the estimate of $\dfrac{u_0}{\sqrt{n}}$, a value of the estimator $S_{\overline{o}}$ obtained in the a particular moment of measurement.

The uncertainty interval of the final measure m of the word error, although approximated because of the adopted estimate procedures, becomes:

$$p = P\left\{\overline{o} - t_p s_{\overline{o}} \le M \le \overline{o} + t_p s_{\overline{o}}\right\}, \tag{9}$$

where t_p is still given by the above recalled Student table.

3. ESTIMATION OF THE WORD ERROR RATE FOR THE SUM OF N SUCCESSIVE OBSERVATION

In this case, the adopted measurement model is that of a final measure M obtained as a result of the sum of r arithmetic means of observations $(\overline{o}_1, \overline{o}_2 \cdots, \overline{o}_r)$ where the generic $\overline{O}_i$ is the mean of n_i observations leading back to random effects (GUM [5] type A) and s random variables $M_1, \cdots, M_j$ affected by systematic effects (GUM [5] type B):

$$M = \sum_{i=1}^{r} \overline{O}_i + \sum_{j=1}^{s} M_j, \tag{10}$$

all the variables belonging to the second part of Eq. (10) are hypothesized as independent. A number of degrees of freedom equal to $\nu_i = n_i - 1$, is given to each $\overline{O}_i$.

Interesting statistical parameters are:

$$\begin{cases} E\left\{\overline{O}_i\right\} = m_{oi}, \ \sqrt{Var\left\{\overline{O}_i\right\}} = u_{oi} \Big/ \sqrt{n_i} & \forall\, i = 1, ..., r \\ E\left\{M_j\right\} = m_j, \ \sqrt{Var\left\{M_j\right\}} = u_j & \forall\, j = 1, ..., s \end{cases}$$

consequently we have:

$$\begin{cases} E\{M\} = \sum_{i=1}^{r} m_{oi} + \sum_{j=1}^{s} m_j \\ \sqrt{Var\{M\}} = u_M = \left[\sum_{i=1}^{r} \dfrac{u_{oi}^2}{n_i} + \sum_{j=1}^{s} u_j^2\right]^{1/2}, \end{cases} \tag{11}$$

In this case, the disposal data are:
the observed values in the estimation of $(m_{o1}, m_{o2}..., m_{or})$
the computed values $(s_{\overline{o}_1}, s_{\overline{o}_2}, \cdots, s_{\overline{or}})$, being $s_{\overline{o}_i} = \sqrt{\frac{\sum_{k=1}^{n_i}(o_{ik}-\overline{o}_i)^2}{n_i(n_i-1)}}$ the generic estimation of $\frac{u_{oi}}{\sqrt{n_i}}$.
the parameters m_j and u_j deduced by a previous type B estimation.

For example, if M_j is included in the $[a, b]$ interval, we can hypothesize an uniform distribution with $m_j = \frac{a+b}{2}$ and $u_j = \frac{b-a}{2\sqrt{3}}$. If this data is available, Eq. (11) can be approximated as follows:

$$\begin{cases} \tilde{E}\{M\} = \sum_{i=1}^{r}\overline{o_i} + \sum_{j=1}^{s} m_j \\ \tilde{u}_M = \left[\sum_{i=1}^{r} s_{\overline{o}i}^2 + \sum_{j=1}^{s} u_j^2\right]^{1/2}. \end{cases} \tag{12}$$

Going back to Eq. (9), if we admit that it is possible to declare, even recalling the central limit theorem, that M follows a normal distribution, we obtain:

$$M = N\left(\tilde{E}\{M\},\ \tilde{u}\{\mathrm{M}\}\right). \tag{13}$$

We accept, in this case, that the variable $T_{\nu_{eff}} = \frac{M - \tilde{E}\{M\}}{\tilde{u}\{M\}}$ follows a Student distribution with the number of degrees of freedom to be determined.

To this end an interesting formula can be taken into account (*Welch-Satterthwaite*) [5], [6] and [7]

$$\nu_{eff} = \frac{\left[\sum_{i=1}^{r} s_{\overline{o}i}^2 + \sum_{j=1}^{s} u_j^2\right]^2}{\left[\sum_{i=1}^{r} \frac{s_{\overline{o}i}^4}{n_i-1} + \sum_{j=1}^{s} \frac{u_j^4}{\nu_j}\right]}, \tag{14}$$

where $\nu_j$is the number of degrees of freedom which can be attributed to u_j^2.

The contributions of the u_j^4 to the denominator of Eq. (14) is sometimes negligible when its number of degrees of freedom tends to infinite.

The final level of uncertainty with a confidence level p is obtained in this case from the following equation

$$P\left\{\tilde{E}\{M\} - t_p\,\tilde{u}_M \leq M \leq \tilde{E}\{M\} + t_p\,\tilde{u}_M\right\} = p, \tag{15}$$

where for a stated confidence level, the corresponding value t_p is determined from the Student table corresponding to the number of degrees of freedom, ν_{eff}, obtained from Eq. (14).

If ν_{eff} is not integer, as is usually the case, then it is necessary to round it off to nearest lower integer.

4. CONCLUSIONS

The aim of this paper is to introduce a new detailed method as a contribution for defining the Word Error Rate measurement. This is made according with GUM [5] and its supplements carried out by BIPM. It has been quantified the uncertainty level with relative confidence level in the two cases of *n* successive observation and of the sum of *n* successive observation using Student and Chi-square distributions.

REFERENCES

1. IEEE Std. 1241, IEEE Standard for terminology an test methods for analog-to-digital converters, 2000.
2. Chiorboli, G.; De Salvo, B.; Franco, G.; Morandi, C.,: "Some thoughts on the word error rate measurement of A/D converters". *Electronics, Circuits and Systems, 1998 IEEE International Conference on Volume 3*, vol.3, 7-10 Sept. 1998, pp.453–456.
3. Box, G. E. P., Hunter, W. G., Hunter, J. S.: *Statistics for experimenters*. John Wiley ,New York, N.Y., 1978.
4. Papoulis, A.: *Probability and Statistics. Englewood Cliffs*. NJ: Prentice-Hall, 1990.
5. Guide to the Expression of Uncertainty in Measurement, first edition (1993), corrected and reprinted (1995), ISO International Organization for Standardization (Geneva, Switzerland).
6. Welch, B. L., (1936), J. R. Stat. Soc. Suppl. 3, 29–48; (1938) *Biometrika* 29, 350-362; (1947), ibid. 34, 28–35.
7. Satterthwaite, F. E. (1941), Psychometrika 6, 309–316; (1946) *Biometrics Bull.* 2(5), 110–114.