Artificial equation errors

Autor(en): **Tichý, Milìk**

Objekttyp: **Article**

Zeitschrift: **IABSE reports of the working commissions = Rapports des commissions de travail AIPC = IVBH Berichte der Arbeitskommissionen**

Band (Jahr): **4 (1969)**

PDF erstellt am: **22.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-5950

VI

Artificial Equation Errors

Erreurs d'équations artificielles

Künstliche Gleichungsfehler

MILÍK TICHÝ
Building Research Institute
Technical University
Prague, Czechoslovakia

Definition of the equation error

In structural analysis and design various prediction formulas are used to describe behaviour of structures, structural elements and sections under different conditions. In most cases, these formulas do not exactly predict the response of the structure, so that the resulting quantities, y^*, deviate less or more from the assumed reality, y, even if exact values or input quantities, g_i, are introduced into the formulas. These deviations are called equation errors; they can be defined either as the differences

$$\Delta = y^* - y$$

or the ratios

$$\lambda = \frac{y^*}{y} .$$

The second definition will be kept in the following text.

Character of the equation errors

Essentially, the equation errors can have two distinctly different characters.

First of all, the equation errors can be caused by unsufficient knowledge of the predicted phenomenon. In most formulas only a certain part of primary quantities influencing the phenomenon (say, g_1 through g_k) is involved, whereas the remaining quantities (g_{k+1} through g_m) are not considered for various rea-

sons. Sometimes even their effect is not known at all.

It is evident that neglecting of only one influence would make the equation error systematical. However, the number of influences and, thus, of primary quantities which are not considered in the formula is frenquently large. Due to this fact and due also to the complexity of phenomena the equation errors acquire random character and, consequently, can be treated by statistical methods.

The equation errors of this type are a necessary but unintentional consequence of our limited possibilities of predicting the phenomena. Therefore, they can be considered as natural.

On the other hand, however, the mathematical model of the phenomena is well known in many cases, but the resulting formulas are too complex. The practical designers often demand their simplification in order to facilitate the design procedures. If, then, an approximate formula giving y^* is found it must always lead to an equation error, λ. It is clear, that this equation error is an intentional, though unwanted consequence of the approximation, and it can be considered as artificial.

Natural equation errors have been already discussed by some authors (Zsutty /1/, Tichý-Vorlíček /2/, Murzewski /3/). Therefore, no special attention will be paid to them in this contribution which, on the other hand, will concentrate on a particular type of the artificial equation error, resulting from simplification of prediction formulas.

Problem formulation

Assume that a function

$$y = f(g_1, g_2, \dots, g_m) \qquad (1)$$

is perfectly defined in the given range, i.e. its magnitude is known for any set of primary quantities g_1 through g_m.

Evidently, function (1) can be replaced by different functions, e.g. by polynomials, Fourier series, etc. However, if a simplification of the original function (1) is desired, the choice of substitutes is rather limited. It appears that for practical purposes an important simplification can be achieved by substituting for (1) the following exponential function

$$y^* = g_1^{x_1} \cdot g_2^{x_2} \cdot g_3^{x_3} \cdots g_n^{x_n} \qquad (2)$$

where x_1 through x_n are constant exponents which are to be found for each separate function and for each set of ranges of primary quantities g_i. By taking the logarithm of Eq. (2) it follows that

$$\log y^* = x_1 \log g_1 + x_2 \log g_2 + \ldots + x_n \log g_n \qquad (3)$$

It is clear that the computation of the investigated quantity is reduced to the summing of logarithms of primary quantities and taking the antilogarithm of the sum. This reduction has a considerable importance by itself, since in most cases the number of mathematical operations involved will be limited to a few. Thus the errors due possibly to the imprecision of the computation itself are to a large extent eliminated and, what is probably often more important, a source of human errors (e.g. the omission of some operation) is lessened. The establishing of exponents x_i is, as it will be shown, in general simple, even if some practical problems must be solved.

Function (2) has already been successfully used by Zsutty /1/ for multiple regression analysis of ultimate strength tests of reinforced concrete sections. The intention of his work was to fit a function to populations of experimental results and to values of primary quantities applying in the ultimate strength which were ascertained in the tests. Equation errors resulting from this are of the natural type.

Method of Solution

The problem of finding the unknown exponents x_i can be solved by using the least squares method for logarithms of y and y^*. The sum of squares of differences between $\log y^*$ and $\log y$ should be minimum, i.e. symbolically.

$$\sum_{j=1}^{M} (\log y_j^* - \log y_j)^2 = \text{minimum} \qquad (4)$$

where M is the number of points for which the difference is found.

Considering the decimal logarithms, it is convenient to put

$$g_1 = 10.0 \tag{5}$$

thus the first factor of the right-hand side of Eq.(2) will be a constant.

Substitute for y^* from Eq. (2) into Eq. (4).

$$\sum_{j=1}^{M} (x_1 \log g_{1j} + x_2 \log g_{2j} + \ldots + x_n \log g_{nj} - \log y_j)^2 = min$$

and differentiate successively by unknowns x_i. In order to minimize the left-hand side of the Eq. (4) the first derivative must be put equal to zero. Hence

$$\sum_{j=1}^{M} (x_1 \log g_{1j} + x_2 \log g_{2j} + \ldots + x_n \log g_{nj} - \log y_j) \log g_{ij} = 0$$

for all $i = 1, 2, \ldots n$.

After rearrangement and taking into account Eq. (5) a system of $n+1$ simultaneous linear equations is obtained:

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = b_1$$

$$a_{12} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = b_2$$

$$\ldots$$

$$a_{1n} x_1 + a_{2n} x_2 + \ldots + a_{nn} x_n = b_n$$

where

$$a_{ik} = \sum_{j=1}^{M} \log g_{ij} \cdot \log g_{kj}$$

$$b_i = \sum_{j=1}^{M} \log g_{ij} \cdot \log y_j$$

It is clear that this system can be easily solved for the unknowns x_i .

If the least squares method is used for substituting a set of experimentally obtained values y_j by a function depending on parameters g_i , the magnitudes of y_j's and g_i's are known from the tests ; they form a discrete population of points. This is schematically shown for y depending on one parameter only, g_2 in Fig.1. However, if applying the method to the substitution of

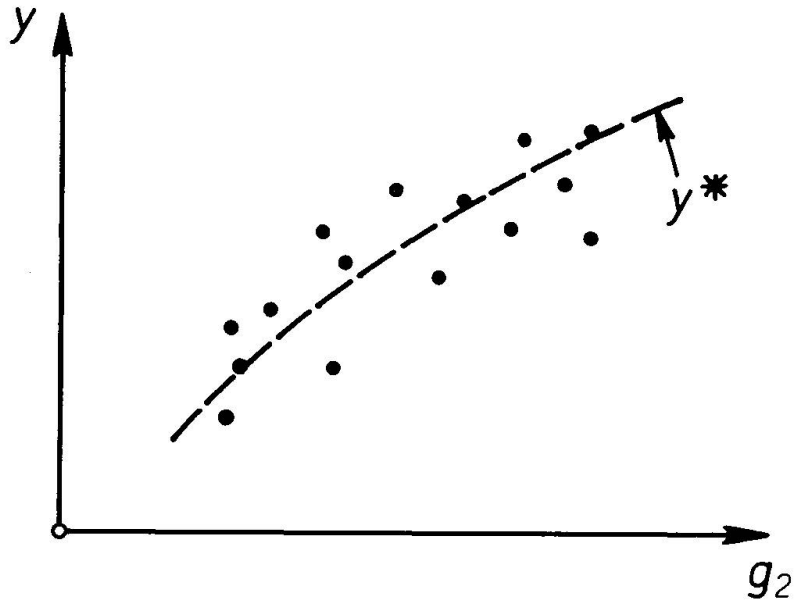

Fig. 1. - Substitute function derived from experimental results.

a defined continuous function, values of y_j and g_i must be artificially generated (Fig. 2). The larger will be the number of generated points, M , the better will be the fit.

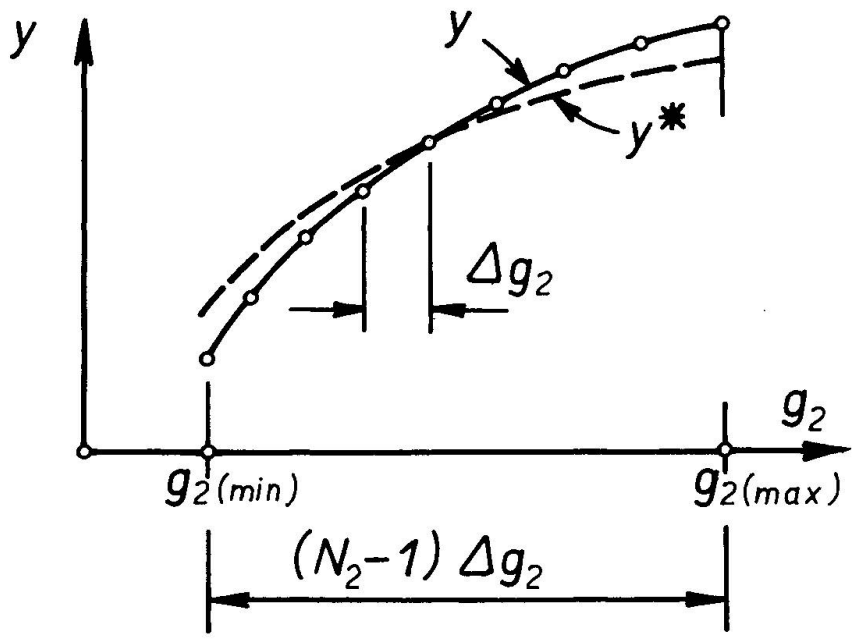

Fig. 2. - Substitue function derived from another defined function.

To generate the population of points, intervals for individual primary quantities g_i must be defined by lower and upper bound values, g_i^I and g_i^{II}, respectively. These intervals are divided into $(N_i - 1)$ divisions, the width of which is Δg_i. Thus, the number of points generated in this way will be given by the product

$$M = \prod_{i=1}^{n} N_i$$

Accepting Eq. (5), no interval is defined for g_1, so that $N_1 = 1$. For all primary quantities the number of divisions may either be equal or it might differ, taking into account the influence of the quantity on the result.

Quasi-randomness of the equation error

It appears from some tentative tests that the distribution of the quantity

$$\lambda = \frac{y^*}{y} \qquad (6)$$

has a surprisingly random character, even if the number of points, M, is relatively small (of the order of 100). This seems to be rather illogical since both functions, y and y^*, are

deterministically given. No explanation of the quasi-random behaviour of λ can be given today. And it cannot be generalised, neither.

The quasi-randomness of the equation error λ suggests the use of the statistical parameters of its distribution as quantities suitable for checking the effectiveness of the substitue function. These parameters should evidently be: the mean $\overline{\lambda}$, the standard deviation s_λ, and also the coefficient of skewness, a_λ, defined by

$$a_\lambda = \frac{\mu_3^c}{s_\lambda^3}$$

where μ_3^c is the third central moment of the statistical distribution of λ. Further statistical parameters are not considered, the above three being quite sufficient for the purpose.

The mean, $\overline{\lambda}$, should be near to unity. Actually, this has proven to be true in all cases investigated to date (for examples of practical application of the method see the author's papers /4/ and /5/). The mean is not greatly influenced by the number of generated points, M.

The standard deviation, s_λ, supplies the information on the spread of λ about the mean. It can be stated, that the less the s_λ, the better the fit. In general, the value of s_λ decreases with the number of points, M; the contribution of individual primary quantities not being uniform, a further decrease of s_λ may be achieved by concentrating the increase in M to some of the quantities only. Since a perfect fit is never possible, the standard deviation converges to some definite value.

A similar importance is attached to the coefficient of skewness, a_λ. In general its value differs from zero, towards negative or positive values. In practical cases values of λ either at the left-hand or at the right-hand tail of the statistical distribution are important, depending upon the nature of the problem. It cannot be said that a value of a_λ near to zero would be the most convenient one - a definite skewness may be often more favourable than a zero skewness.

If computing the statistical parameters of the population

of values λ obtained for those M generated points which were used in deriving the unknown exponents x_i, incorrect conclusions might be drawn. Evidently, in this way the quality of the fit is checked just by those points from which x_i's has been established. Therefore, it is necessary to check the fit also for points which are removed as far as possible from the M points of the original population. Such points lie just between the original ones (for one-dimensional function this is schematically shown in Fig.3). A new, control population of values y and y^* is generated for the control system of points and

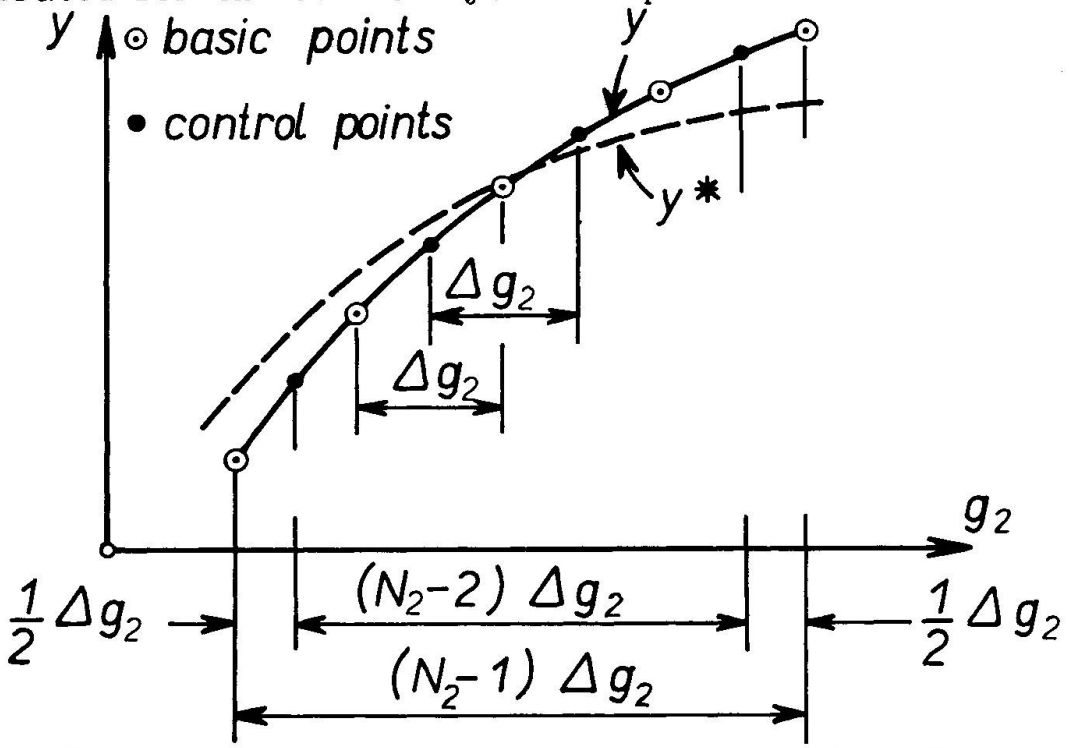

Fig. 3. - Basic and control system of values (y, g_2) for function y depending on one primary quantity.

again the statistical parameters of the equation error, λ, are found. It is obvious that the number of points of the control population will be smaller that M, being equal to

$$m = \prod_{i=1}^{n} (N_i - 1)$$

For the assessment of the new set of statistical parameters the same is valid as for the original population.

Partial safety factor

Modern structural code formats as introduced by the European Concrete Committee, CEB, International Building Council, CIB, and others use to cover the approximations of design assumptions the partial safety factor γ_a belonging to a wider family of factors γ. This factor is frequently introduced only in conceptual terms, or sometimes, directly by means of an empirical value; yet, no method of establishing γ_a has been so far presented.

Now, using the above quasi-randomness of the equation error, λ, the partial safety factor γ_a can be defined. The procedure is outlined as follows: Computing the value of y^* for a given set of g_i's it is known that in comparison with the exact value of y the result is charged by equation error λ. However, the magnitude of λ is not known. On the other hand it is known that the quasi-randomness of λ is described by the statistical parameters established for the whole population of points investigated. Assuming now a convenient statistical distribution for the description of the random behaviour of λ (Fig. 4) the adequate quantile, either λ_{min} or λ_{max}, can be found for a chosen probability $P(\lambda < \lambda_{min})$ or $P(\lambda > \lambda_{max})$.

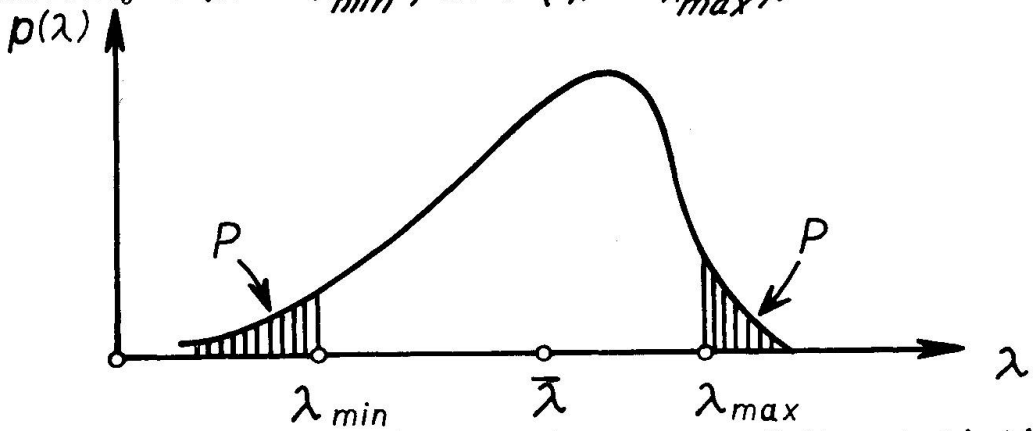

Fig. 4. - Probability density curve of the statistical distribution of λ.

The values of the quantiles are given by

$$\lambda_{min} = \bar{\lambda} + t_{min}\, s_\lambda \tag{7}$$

$$\lambda_{max} = \bar{\lambda} + t_{max}\, s_\lambda \tag{8}$$

where t_{min} (with negative sign) or t_{max} are values of standardized random variable (with zero mean and unit variance) depending on the type of the selected distribution and its statistical parameters, and on the probability $P(\lambda < \lambda_{min})$ or $P(\lambda > \lambda_{max})$, respectively.

At the present state of knowledge it is not possible to say in general which type of statistical distribution should be used. In the cases investigated until now (see /4/ and /5/) log-normal distribution proved to be adequate for modelling the quasi-randomness of λ (it must be mentioned here that the log-normal distribution can have both positive or negative skewness; this fact is not commonly known). In other cases, of course, other types od statistical distribution might be applied. It is evident that the optimum type would depend upon the type of the substitute function y^*.

Similarly, no definite answer, based on some theoretical analysis can be given with regard to the probability $P(\lambda < \lambda_{min})$, $P(\lambda > \lambda_{max})$. The problem must be solved within the whole context of statistical design. However, it may be tentatively said that a reasonable value of P would be 10^{-2} for cases of minor importance (e.g. checking the stiffness of a current beam) and about 10^{-3} in more important cases (ultimate strength, etc.). It must be stressed here that the probability P has nothing to do with the probability of failure P_U, since it represents entirely different statistical phenomenon. The use of the method is not restricted to a probabilistic code format only, it can be used also with the classical codes.

Assume, now, for instance, a quantity y the over-estimate of which in comparison with its real value is on the unsafe side in the design (e.g. the ultimate moment of a section). After finding the substitute value y^* it is obviously not known if its deviation from y is at the left or right-hand tail of the distribution of λ. To be on the conservative side the worst must be anticipated and, therefore, λ_{max} must be considered in the design, i.e. in order to obtain the design value y^d, y^* must be divided by λ_{max}, i.e.:

$$y^d = \frac{1}{\lambda_{max}} \cdot y^*$$

The same procedure is used for cases where negative deviations are unsafe:

$$y^d = \frac{1}{\lambda_{min}} \cdot y^*$$

Consequently, the partial safety factor γ_a covering the equation error is defined as

$$\gamma_a = \frac{1}{\lambda_{max}} \tag{9}$$

or

$$\gamma_a = \frac{1}{\lambda_{min}} \tag{10}$$

It is clear that for cases where λ_{min} is to be used negative skewness of the distribution will be more favourable than a positive one; the opposite is true for λ_{max} .

The use of the factor γ_a can be seen from the following example:

The condition of safety of a reinforced concrete section for the limit state of failure is e.g.

$$1.5 M_D + 1.8 M_L \leqq \phi M_U \tag{11}$$

where M_D , M_L are moments produced in the section by dead load and live load respectively, M_U is the ultimate moment of the section calculated by means of an exact formula and ϕ is the capacity reduction factor covering the random behaviour of the materials, dimensions etc. (actually, ϕ is again a partial safety factor).

Using now a substitute formula for the ultimate moment, value M_U^* is calculated and the condition of safety (11) must be changed to

$$1.5 M_D + 1.8 M_L \leqq \gamma_a \phi M_U^*$$

where evidently in thus case γ_a is defined by $1/\lambda_{max}$, i.e. $\gamma_a < 1$, since higher design ultimate moment, M^d, would give an unsafe result. For further examples see again /4/ and /5/.

Acknowledgements

The present paper has been prepared in the highly creative atmosphere of the University of Waterloo, Waterloo, Canada, during the author's visiting appointment. Thanks are extended to the University for the availability of its excellent computing facilities, particularly of the IBM System 360 Model 75.

- - o o - -

References

1. - Zsutty, T. : Error statistics for reinforced concrete. Proceedings of ASCE, Journal of Structural Division, 1963/3 ST6.

2. - Tichý, M. - Vorlíček, M.: Statistical Theory of Concrete Structures (a book to be published in English by ACADEMIA, Prague, Czechoslovakia).

3. - Murzewski, J. - Sowa, A.: Assessment of error in computing load-carrying capacity of reinforced concrete member (in Polish). Archivum inżynierii lądowej, 1969/1-2.

4. - Tichý, M.: Simplification of design formulas. Acta Technica ČSAV (Prague), 1969/4.

5. Tichý M.: Approximate computation of the sectional stiffness of reinforced concrete beams (in Czech). Stavebnícky časopis, 1969/10.

SUMMARY

When known deterministic formulas are approximated by more simple formulas artificial equation errors are encountered. The deviation between the results obtained by both formulas can be mathematically treated. It has been found in some cases that the behaviour of the artificial equation error is quasi-random, so that it can be described by a suitable statistical distribution. This fact has a significance for the practical design, since partial safety factors can be mathematically derived.

RESUME

Si des formules déterministiques connues sont approximées par des formules simplificatives, les erreurs d'équation artificielles entrent dans le calcul. Les dérivations entre les résultats donnés par les deux formules (originale et simplifiée) peuvent être traitées mathématiquement.

Il a été trouvé en quelques cas étudiés que l'erreur d'équation artificielle est quasi-aléatoire et qu'on le peut décrire par une distribution statistique convenable. Ce fait a une importance pour le calcul pratique: le coefficient partiel de sécurité de calcul approximatif peut être dérivé par des méthodes mathématiques.

ZUSAMMENFASSUNG

Wenn man genaue durch vereinfachte Funktionen darstellt, entsteht ein künstlich geschaffener Gleichungsfehler (wie er in der klassischen linearen Regressionsrechnung als im Quadrat zu minimierende Abweichung auftritt). Die Abweichung der durch die beiden Funktionen entstehenden Ergebnisse kann berechnet werden. Der künstliche Gleichungsfehler verhält sich sozusagen zufällig, so dass er durch eine Dichtefunktion dargestellt werden kann. Dieser Umstand hat einen für die Praxis unübersehbaren Vorteil, sintemal Teilsicherheitsbeiwerte mathematisch hergeleitet werden können.

Leere Seite
Blank page
Page vide