INACCURACIES OF MEASURING METHODS AND THEIR INFLUENCE ON THE REGRESSION FUNCTION†

W. SAUER
Technical University Dresden, DDR

(Received November 11, 1982; in final form June 6, 1983)

The quality parameters of electronic components and devices usually depend on the parameters of the materials. In many cases one does not know the theoretical relationship between the parameters, and therefore one makes technological experiments and measures the values of the parameters. Usually it is necessary to take several measuring points and calculate from this the unknown relationship between the parameters.

The simplest equation that one can use is the linear function. In this case the theoretical probability density function is a Gaussian function. Otherwise it is necessary to assume that the linear function is an approximation.

When the measuring process has an inaccuracy, then one can show that the increase of the linear function is smaller and it is necessary to estimate a factor of correction to calculate the theoretical or exact relationship.

1. THE REGRESSION FUNCTION

The quality characteristics of electronic components are largely dependent on the material characteristics, among which also the geometric dimensions will be classified by us, in general show variations in their values, we observe also a variation in the values of the quality characteristics.

Normally, between the different parameters and characteristics there exist relations of a physical or chemical type, which are not precisely known, due to a large number of influencing factors. For instance, it has been estimated that the yield in the manufacturing of integrated circuits is influenced by 500-600 factors.

In production engineering the required relationships are determined empirically. Let us consider here an influencing quantity X and a quality characteristic Y. In the Figure 1, as an example, the reverse voltage is represented as a function of the temperature during rinsing. The points are determined by measuring. The science of engineering now requires methods for constructing a relation from this point distribution.

One possibility consists in determining the relative number of points in a unit area and estimating from it the so-called density function f(x, y). Of course, this is very time-consuming, but it would reflect the relationship quite well.

Since with two-dimensional functions the relation between X and Y is not obtained directly, one calculates with the so-called regression function.

It represents the conditional expectation value

\[ E \{ Y|X = x \} = \frac{\int f(x, y) \, dy}{\int f(x, y) \, dy} = \mu_y(x). \] (1)

The expression in the denominator is called boundary density f_y(x):

\[ f_y(x) = \int f(x, y) \, dy. \] (2)

†Paper originally presented at the 5th International Spring Seminar on Electrotechnology held at Prenet, Czechoslovakia, 1–4 June, 1982.
In most simple cases, "Eq. (1)" is a straight line (Figure 1) (regression line)

\[ \mu_y(x) = \mu_y + \alpha \cdot (x - \mu_x), \]  

(3)

where

- \( \mu_x = \) expectation value of the boundary distribution \( f_x(x) \)
- \( \mu_y = \) expectation value of the boundary distribution \( f_y(y) \)
- \( \alpha = \) coefficient of slope.

As is known, the coefficient of slope \( \alpha \) is calculated from the covariance and the variance of the boundary density \( f_y(x) \) as

\[ \alpha = \frac{\sigma_{xy}}{\sigma_x^2}. \]  

(4)

Thus, the regression line always passes the point of intersection of the expectation values.

In practice, all coefficients are calculated from the measuring points, and the empirical straight line (empirical regression line)

\[ \bar{y}(x) = \bar{y} + a \cdot (x - \bar{x}) \]  

(5)

is obtained, with the mean values
MEASURING METHODS AND THE REGRESSION FUNCTION

\[
\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n; \quad \bar{y} = \frac{1}{N} \sum_{n=1}^{N} y_n
\]

(\(N\) being the total number of measuring points).

and the empirical coefficient of slope \(a\)

\[
a = \frac{s_{xy}}{s_x^2}
\]

\(s_x^2\) is the empirical variance

\[
s_x^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x - \bar{x})^2
\]

and \(s_{xy}\) is the empirical covariance

\[
s_{xy} = \frac{1}{N-1} \sum_{n=1}^{N} (x - \bar{x})(y - \bar{y})
\]

The straight line in Figure 1 can be calculated by means of this. Of course, it may deviate slightly from the exact straight line. In practice, however, the relation resulting from this measuring process is always used for further operations.

2. THE INACCURACY OF THE MEASURING PROCESS

By the measuring process the value \(u\) is determined for the true value \(x\), and the value \(v\) for the true value \(y\). Of course, \(x\) and \(u\) or \(y\) and \(v\) may agree, but they differ from each other, in general.

Here, also, a two-dimensional density \(g(u, v; x, y)\) is introduced, which, however, is still dependent on \(x\) and \(y\).

In many cases holds

\[
g(u, v; x, y) = g(u-x, v-y) = g(x', y'),
\]

i.e., density depends only on the difference between the read value and the true one. By this, the density can be interpreted as a two-dimensional density for the measuring process as well as for the production engineering process, and the same considerations hold with respect to a regression function.

For instance, for the "pure" measuring process alone there can be given a regression line

\[
\bar{y}(x) = \bar{y} + a'(x-x')
\]

with the slope

\[
a' = \frac{s'_{xy}}{s_x'^2}
\]

where \(s'_{xy}\) is the covariance of the measuring processes and \(s_x'^2\) is the variance of the boundary distribution. \(x'\) and \(y'\) are the systematic measuring errors with respect to \(x\) and \(y\) (expectation values of the boundary distributions). Frequently, the measurements for \(x\) and \(y\) are stochastically independent, i.e., \(s'_{xy} = 0\).
Instead of the variables \( x' \) and \( y' \), in mathematical operations frequently \( x \) and \( y \) are also used — this being only a question of designation.

3. THE PRODUCTION ENGINEERING PROCESS AND THE INACCURACY OF MEASURING

Since the measuring process does not have any influence on the stochastic properties of the production engineering processes, and vice versa, the total process is determined by convolution:

\[
h(x, y) = f(x, y) * g(x, y)
\]

\[
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x', y') \cdot g(x-x', y-y') \, dx' \, dy'.
\]

Then, if the parameters of the total process are labelled with an asterisk (*) and the parameters of the measuring process with a prime, one obtains:

\[
\bar{x}^* = \bar{x} + \bar{x}'
\]

\[
\bar{y}^* = \bar{y} + \bar{y}'
\]

\[
s_{x^2}^* = s_{x^2} + s'_{x^2}
\]

\[
s_{y^2}^* = s_{y^2} + s'_{y^2}
\]

\[
s_{xy}^* = s_{xy} + s'_{xy}
\]

as well as for the slope

\[
a^* = \frac{s_{xy} + s'_{xy}}{s_{x^2} + s'_{x^2}}.
\]

Thus, for the total process the regression line is obtained as

\[
\bar{y}^*(x) = \bar{y} + \bar{y}' + \frac{s_{xy} + s'_{xy}}{s_{x^2} + s'_{x^2}} \cdot (x - \bar{x} - \bar{x}')
\]

In practice, for most cases hold \( x' = 0, y' = 0 \) and \( s'_{xy} = 0 \). Then, the slope of the regression line is reduced by the inaccuracy of the measuring process:

\[
\bar{y}^*(x) = \bar{y} + \frac{s_{xy}}{s_{x^2} + s'_{x^2}} (x - \bar{x}).
\]

Since in practice there are the true process points not known, but always only the real measuring points, for the case represented above the regression line

\[
\bar{y}^*(x) = \bar{y}^* + \frac{s_{xy}}{s_{x^2}^*} (x - \bar{x}^*)
\]

is determined by measurement and the relationship of the pure production engineering process is calculated from it, the stochastic measuring error \( s_{x^2}^* \) being known.
y(x) = y* + \frac{s_{xy}}{s_{x^2} - s_{y^2}} (x - x*). \quad (18)

In addition, from the degree of determinacy of the total process

\[
\rho^{*2} = \frac{s_{xy}^2}{s_{x^2}^2 - s_{y^2}^2} \quad (19)
\]

by the correction by means of the two measuring errors \( s_{x^2} \) and \( s_{y^2} \) one can calculate the degree of determinacy for the pure production engineering process:

\[
\rho^2 = \frac{s_{xy}^2}{(s_{x^2}^2 - s_{x^2}^2) - (s_{y^2}^2 - s_{y^2}^2)}. \quad (20)
\]

Thus, by the measurement also the correlation coefficient and thus the degree of determinacy \( \rho^2 \) is reduced. Of course, the test for stochastic independence must be carried out with \( r \) and not with \( \rho \).

4. GENERAL RELATIONSHIPS

Also, with nonlinear regression functions the influence of the measuring process can be predicted. For this, one requires the boundary density \( g_y(x) \) of the measuring process, or the measuring density \( g(x) \) in the case of a stochastic independence. Then,

\[
\mu_y(x) = \frac{[\mu_y(x) * f_y(x)] * g_y(x)}{f_y(x) * g_y(x)}; \quad (21)
\]

Of course, this expression is not easy to handle.

5. APPLICATIONS

For the first time, these methods were used to analyse production engineering processes of microelectronics. Here, however, the general validity existing for any measuring process should be pointed out, also when the inaccuracy of measuring may be neglected now and then.

REFERENCES

1. Fisz, M. "Wahrscheinlichkeitsrechnung und mathematische Statistik" Berlin VEB Deutscher Verlag der Wissenschaften (1962).
2. Doetsch, G. "Anleitung zum praktischen Gebrauch der Laplace-Transformation" München: R. Oldenbourg Verlag (1956).
3. Sauer, W. "Ein Additionstheorem für Regressionsfunktionen und seine Anwendung in der technologischen Prozeßanalyse" Wiss. Zeitschrift der TU Dresden 32 (1983) H.3 S39–43
Submit your manuscripts at http://www.hindawi.com