ON THE DETECTION LIMIT

CANDIN LITEANU and ION RÎCĂ

*Department of Analytical Chemistry, University of Cluj-Napoca,
Cluj-Napoca, Romania*

ABSTRACT

In order to define the detection limit the two-step model is used, based on a false detection probability P_{10} (error of the first kind) to which there is a corresponding detection level as well as a proper detection probability P_{11}. Given two probabilities P_{10} and P_{11} one may associate them with the detection level y_k as well as with the detection limit $\bar{y}_d$, respectively $\bar{c}_d$, two values of the entropy H_k, respectively H_d. In order to estimate the detection limit two procedures are given: the procedure based on the calibration function $\bar{y} = f(c)$ obtained for values $c < \bar{c}_d$ and the frequentometrical procedure based on the calibration function $P_{11} = f(c)$, where $c < \bar{c}_d$. In both cases a confidence interval is associated with the detection limit. Finally, there is a description of the manner in which three tests are applied in order to check the hypothesis concerning the detection of a component in the sample under investigation: a parametrical (Student), a non-parametrical (Wilcoxon) and a sequential (Wald) one.

1. DEFINITION OF DETECTION LIMIT

Like any experimental magnitude, the analytical signal also depends on a very large number of variables:

$$y = f(x_1, x_2, \ldots, x_n) \qquad (n \to \infty) \tag{1}$$

and since all variables x have a random character, the analytical signal has obviously a random character, i.e. it always includes a certain uncertainty. Its distribution will be normal (Gaussian) when relation (1) is described by the Liapunov–Lindeberg central limit theorem.

The uncertainty (deviation of the measured values from the actual ones) which accompanies any analytical measurement has therefore a very complex nature and always includes the background noise, whose fluctuations in most cases have a Gaussian distribution.

The perturbation of the analytical signal occurs in all phases of an analytical system, as seen in *Figure 1*.

We shall consider below only the perturbations arising in the measurement step, which obviously also include perturbations occurring in the previous steps (sampling and physico-chemical treatments).

An analytical system, as a result of the examination of a material, must solve the following problems: (a) the problem of detection (qualitative analysis) and (b) the problem of determination (quantitative analysis).
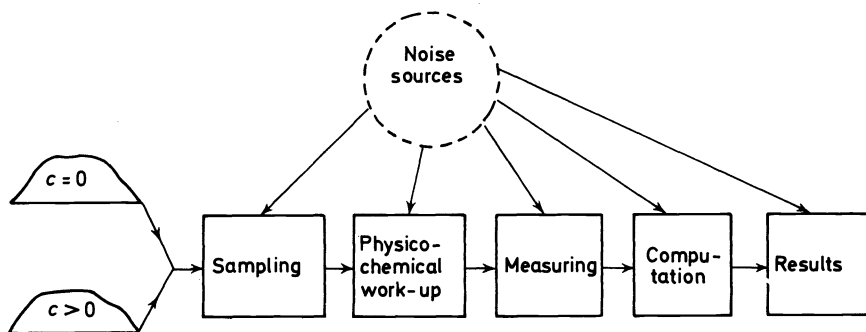
535

Figure 1. General scheme of an analytical system

As seen in *Figure 1*, in the case of an analytical system, the material to be analysed is at the input and the result of the analysis is at the output of the system.

In the case of the detection, the material at the input of the system may be present in two states: with the component to be identified ($c > 0$) and without the component to be identified ($c = 0$).

Figure 2 shows the relation between the average value of the fluctuations of the background $\bar{y}_0$ and the average value of the analytical signal $\bar{y}$, for a given concentration value[1]. One can observe that in some cases, due to interferences, an assumption may be made as to whether the signal belongs to the background or to concentration.

The relationship between the background noise and the analytical signal in the frame of the problem of the definition and evaluation of the detection limit initially was studied by Kaiser[1] and subsequently in other papers[2-20].

Figure 3 presents the graphical model for a linear dependence between the concentration c and the analytical signal y, in the presence of a Gaussian noise.

In this case, one may write:

$$y \Rightarrow N(\bar{y} = \bar{y}_0 + bc; \sigma_y^2) \tag{2}$$

i.e. the analytical signal y is normally distributed with a mean $\bar{y}$ and dispersion σ_y^2; $\bar{y}_0$ is the mean of background fluctuations.

The acceptance of such a model to discuss the problem of the analytical detection is justified by the following: (1) in most cases, for low concentrations, the relationship between the analytical signal and concentration is linear; (2) the values of the analytical signal in most cases have a Gaussian distribution; (3) the dispersion of the analytical signal σ_y^2, in general, changes little for the narrow variations of concentration.

An examination of the model in *Figure 3* shows that as the concentration increases, the probability field for the analytical signal will have fewer and fewer common elements with the probability field for the background signal ($c = 0$). Only at higher concentrations of a certain value will the two probability fields become practically separated and therefore the detection of the component in question will raise no problems.
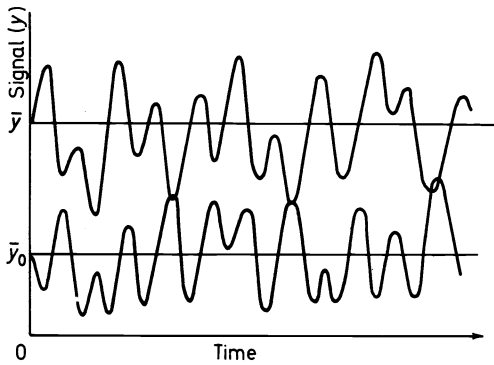
536

Figure 2. Fluctuations of the background and the analytical signal, respectively, in the case of a given concentration value[1]
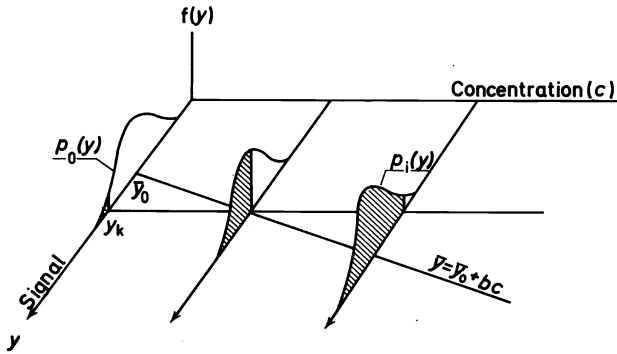

Figure 3. Graphical model of the variation of the signal distribution for a linear dependence between concentration c and analytical signal y in the presence of a Gaussian noise
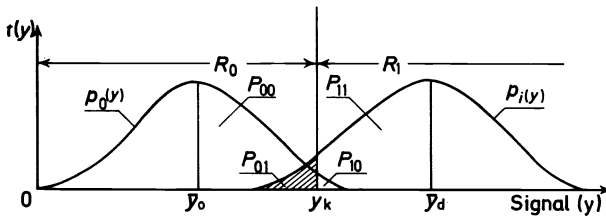

Figure 4. Two-step statistical model for defining the detection limit

Figure 4 shows the relation between the fluctuation fields of the background signal ($c = 0$) with the probability density $p_0(y)$ and that of the analytical signal $p_i(y)$ corresponding to concentration c_i.

Considering the model presented in *Figure 4*, as a result of a measurement,

537

we must choose, based on a criterion, between two hypotheses: H_0 (the signal belongs to the probability density $p_0(y)$) and H_1 (the signal belongs to the probability density $p_i(y)$). This criterion aims to divide the values of y into two regions R_0 and R_1, so that for a value of a measurement situated in the region R_1 we may accept the hypothesis H_1 ($y > y_k$), i.e. the component is present.

Consequently, all the values of the analytical signal situated in the region R_1 permit us to accept, with a certain probability, the hypothesis present, from the two possibilities: present or absent respectively. In this manner, all the values $y > y_k$ provide an amount of information of one bit.

The value of the analytical signal y_k which divides the value field of signal y into two regions is called *critical level*[18], *decision level*[18, 20] or, more suitably, *detection level*. Under these conditions, the following probabilities may be distinguished, which characterize the detection process:

1. Probability of choosing hypothesis H_0 when it is true:

$$P_{00} = \int_{-\infty}^{y_k} p_0(y)\, dy \tag{3}$$

2. Probability of choosing hypothesis H_1 when H_0 is the true one (error of the first kind):

$$P_{10} = \int_{y_k}^{\infty} p_0(y)\, dy \tag{4}$$

3. Probability of choosing hypothesis H_0 when H_1 is the true one (error of the second kind):

$$P_{01} = \int_{-\infty}^{y_k} p_i(y)\, dy \tag{5}$$

4. Probability of choosing hypothesis H_1 when it is true:

$$P_{11} = \int_{y_k}^{\infty} p_i(y)\, dy \tag{6}$$

Relations

$$\left.\begin{array}{c} P_{00} + P_{10} = 1 \\ P_{01} + P_{11} = 1 \end{array}\right\} \tag{7}$$

are evident.

From these last relations it results that for the characterization of the identification process two probabilities are sufficient (one each for the two fields R_0 and R_1).

In order to establish the detection level y_k, or indeed the probability P_{10} of false detection (error of the first kind), one may use several criteria: the criterion of minimum risk, the criterion of maximum verisimilitude, the criterion of the ideal observer, the Neyman–Pearson criterion, etc.[21, 22].

The Neyman–Pearson criterion calculates the detection level based on a previously imposed false detection probability P_{10}.

By normalization of the normal distribution law, i.e. if

$$z_k = (y_k - \bar{y}_0)/\sigma_{y_o} \qquad (8)$$

(σ_{y_o}—standard deviation of background fluctuations), one obtains

$$P_{10} = (1/\sqrt{2\pi}) \int_{y_k}^{\infty} \exp\left(-\tfrac{1}{2}z^2\right) dz \qquad (9)$$

and thus, since $P_{10} < 0.5$, one obtains

$$\Phi(z_k) = 0.5 - P_{10} \qquad (10)$$

Then, using tables[23] with the Laplace function values $\Phi(z)$, one obtains the z_k value, i.e.

$$y_k = \bar{y}_0 + \sigma_{y_o}z_k \qquad (11)$$

namely, the value of the signal corresponding to the detection level.

Then it is necessary to consider a proper identification probability P_{11}, corresponding to the correct hypothesis H_1.

Since $P_{11} > 0.5$,

$$\Phi(z_d) = P_{11} - 0.5 \qquad (12)$$

and from the tables with the Laplace function values one obtains:

$$z_d = (\bar{y}_d - y_k)/\sigma_{y_o} \qquad (13)$$

whence

$$\bar{y}_d = y_k + \sigma_{y_o}z_d \qquad (14)$$

In expression (14), σ_{y_o} has the same value as in expression (11), which is justified by the fact that the concentration field corresponding to interval $\bar{y}_0 \ldots \bar{y}_d$ is low (see p. 536).

From equations (11) and (14),

$$\bar{y}_d = \bar{y}_0 + (z_k + z_d)\sigma_{y_o} = \bar{y}_0 + k\sigma_{y_o} \qquad (15)$$

Since the value of signal y always includes two components, i.e. that corresponding to the background and that due to the concentration, the *net analytical signal* will be given by the difference $\bar{y}_d - \bar{y}_0$, i.e.

$$\bar{y}_d - \bar{y}_0 = k\sigma_{y_o} \qquad (15')$$

In conclusion, the coefficient $k = (\bar{y}_d - \bar{y}_0)/\sigma_{y_o}$ is given by the ratio between the net analytical signal and the background noise whose measure is the standard deviation σ_{y_o}.

Considering, for instance, the detection level y_k for which the false identification probability of the first kind $P_{10} = 0.001$ and also the proper identification probability $P_{11} = 0.998$, it is found that $z_k = 3.1$ and $z_d = 2.9$, so that for the detection limit under such conditions one obtains

$$\bar{y}_d = \bar{y}_0 + 6\sigma_{y_o} \qquad (16)$$

In conclusion, *the detection limit may be defined by that value of concentration $\bar{c}_d$, for which, with regard to the detection level y_k, that is admitting a*

certain false detection probability value P_{10}, to the analytical signal $\bar{y}_d$ corresponds a correct detection probability P_{11}.

Figure 5 shows the relationship between the mean of the background noise fluctuations and the mean of the analytical signal fluctuations corresponding to the detection limit in the frame of the two-step model.
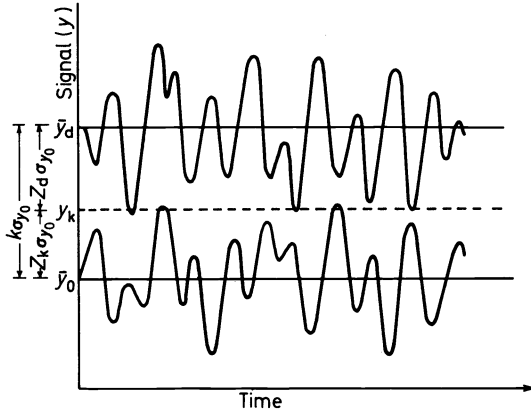

Figure 5. Two-step graphical model for the relationship between the background signal (background noise) and the analytical signal corresponding to the detection limit

Returning to the two-step detection model presented in Figure 4, we see that two events are attached to the density of probability for the background signal $p_0(y)$ in relation to the detection level y_k: $y < y_k$ and $y > y_k$, for which $P(y < y_k) = P_{00}$ and, respectively $P(y > y_k) = P_{10}$, so that $P_{00} + P_{10} = 1$, and, hence, the two events form a complete system of events, for which the entropy will be

$$H(P_{00}, P_{10}) = - P_{00} \log P_{00} - P_{10} \log P_{10} = H_k \tag{17}$$

and since $P_{00} = 1 - P_{10}$ (see equation 7), it results:

$$H(P_{00}, P_{10}) = - (1 - P_{10}) \log (1 - P_{10}) - P_{10} \log P_{10} = H_k \tag{18}$$

Thus, for $P_{10} = 0.001$, $H_k = 0.0114$ bit.

Figure 6(a) represents relation (18) which expresses the dependence of the false detection entropy H_k on the value of the signal corresponding to the detection level defined by means of the Neyman–Pearson criterion.

In conclusion, the detection level y_k may be defined by that signal value for which, considering a certain false detection probability P_{10}, the entropy of events $y < y_k$ and $y > y_k$ has a certain value H_k.

In relation to the detection level y_k, the density of probability for the analytical signal $p_i(y)$ also delimits two events $y < y_k$ and $y > y_k$ for which $P(y < y_k) + P(y > y_k) = P_{01} + P_{11} = 1$, i.e. a complete system of events, so that the expression of the entropy H_d will be

$$H(P_{01}, P_{11}) = - (1 - P_{11}) \log (1 - P_{11}) - P_{11} \log P_{11} = H_d \tag{19}$$

Thus, for $P_{11} = 0.998$, $H_d = 0.0208$ bit. *Figure 6(b)* presents relation (19) which expresses the dependence between the correct detection entropy H_d and the values of the analytical signal.

In conclusion, the *detection limit may be defined by that concentration value $\bar{c}_d$ to which, with regard to a certain level of the false detection entropy $H(P_{00}, P_{10})$ evaluated on the basis of the detection level y_k, there corresponds a certain value of the proper detection entropy $H(P_{01}, P_{11})$.*
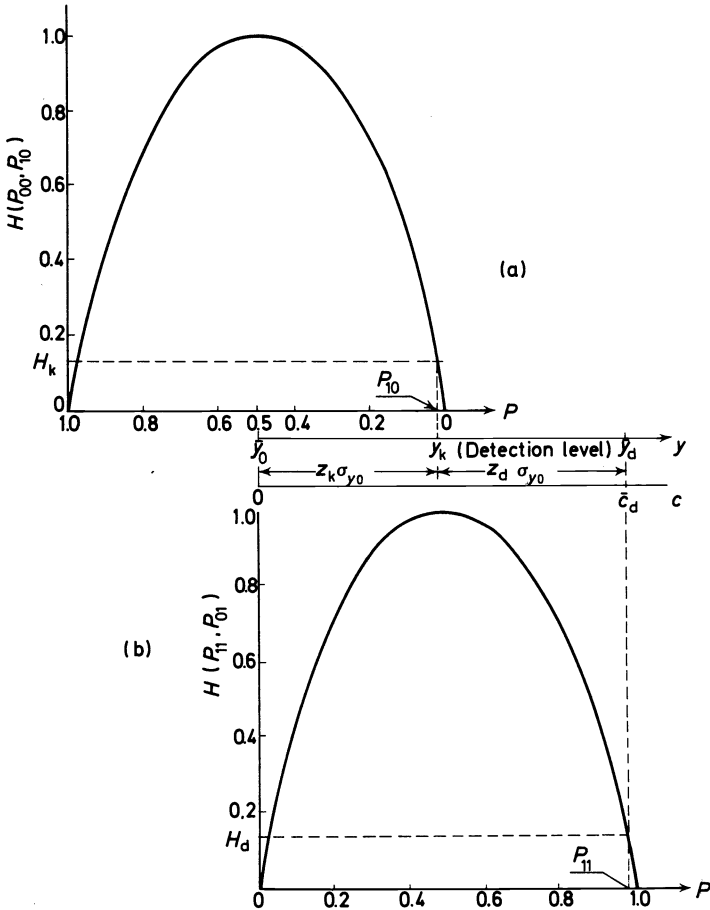

Figure 6. Two-step information model for defining the detection limit

2. ESTIMATION OF THE DETECTION LIMIT

We shall give below two procedures for estimating the detection limit:
(1) the procedure based on the calibration function $\bar{y} = f(c)$ obtained for

values $c < \bar{c}_d$ and (2) the frequentometrical procedure based on the calibration function $P_{11} = f(c)$, where $c < \bar{c}_d$.

Estimation of the detection limit based on the calibration function $\bar{y} = f(c)$

In order to calculate the detection limit by means of formula (15), the value of the analytical signal $\bar{y}_d$ must be converted into concentration. To do this, it is necessary to know the correlation $\bar{y} = f(c)$ in a concentration range as close as possible to the detection limit. Analytical signal $\bar{y}$ for a given concentration in the range $c < \bar{c}_d$, is the mean value of N repetitions (see *Table 1*). For this purpose, in the concentration range higher than the detection limit, using the pairs of values $\bar{y}_i, c_i$, one calculates (least squares method) the calibration function

$$\bar{y} = a + bc = \bar{y}_0 + bc \qquad (20)$$

Even in the case of a non-linear correlation, one may eventually obtain a linear correlation in the form of equation (20), by appropriate procedures (change of variables). In equation (20) the intercept $\bar{y}_0$ represents exactly the estimated mean of the background oscillations ($c = 0$). As seen in *Figure 7*, to the calibration function (20) is attached a dispersion (confidence) band, limited by two arcs of the hyperbola whose equations are[24]

$$\bar{y} = \bar{y}_i \pm \frac{t_{(P,n-2)}(SPD)^{\frac{1}{2}}[S_{cc} + n(c_i - \bar{c})^2]^{\frac{1}{2}}}{[n(n-2)S_{cc}]^{\frac{1}{2}}} \qquad (21)$$

where $\bar{y}_i$ is the value of the signal corresponding to concentration c_i in the calibration function (20); $t_{(P,n-2)}$ is the Student variable (test) for probability P and $n - 2$ degrees of fredom; n is the number of pairs of values $\bar{y}_i, c_i$ used for the calibration function; and $\bar{c} = (\sum_{i=0}^{n} c_i)/n$ is the mean of concentration values used for obtaining the calibration function (20)

$SPD = S_{yy} - bS_{cy}$ is the sum of squares of deviations from function (20):

$$S_{yy} = \sum_{i=1}^{n} \bar{y}_i^2 - (\sum_{i=1}^{n} \bar{y}_i)^2/n$$

$$S_{cy} = \sum_{i=1}^{n} c_i \bar{y}_i - \sum_{i=1}^{n} c_i \sum_{i=1}^{n} \bar{y}_i/n$$

$b = \dfrac{S_{cy}}{S_{cc}}$ is the slope of function (20)

$$S_{cc} = \sum_{i=1}^{n} c_i^2 - (\sum_{i=1}^{n} c_i)^2/n$$

$\bar{y}_0 = \bar{y}^* - b\bar{c}$ is the intercept, i.e. the estimated value for the mean of background fluctuations, in which $\bar{y}^* = (\sum_{i=1}^{n} \bar{y}_i)/n$ is the mean of all values of signals used for obtaining the calibration function (20).

Then the analytical signal $\bar{y}_d$ corresponding to the detection limit will be calculated by equation (15) written in the form

$$\bar{y}_d = \bar{y}_0 + ks_{\bar{y}} \qquad (22)$$

i.e. in which σ_{y_o} was replaced by its estimate $s_{\bar{y}}$ obtained from the calibra-
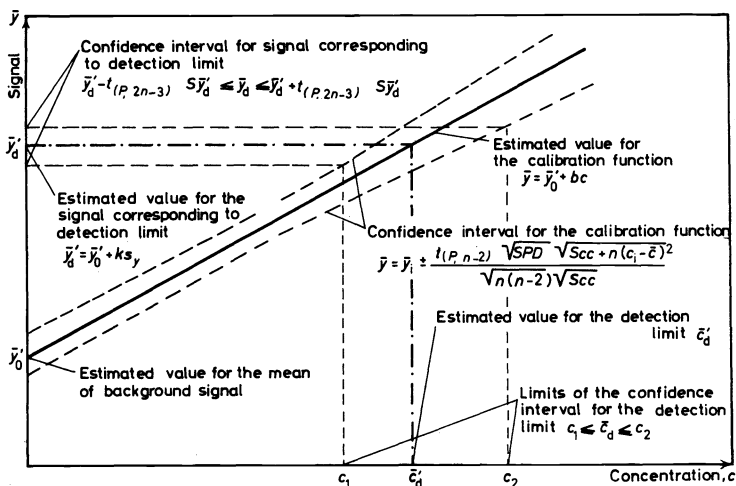
Figure 7. Evaluation of the detection limit from the calibration function

tion data (function 20), a replacement justified by the fact that, as has been shown, one works in a concentration field close to the detection limit. In this case, obviously, $\bar{y}_0$ and also $\bar{y}_d$ will be estimated magnitudes as well. In order to determine the confidence interval of the detection limit, we must first calculate the confidence interval of the analytical signal. To this end, one considers the random variable

$$t = (\bar{y}_d - \bar{y}_d')/s_{\bar{y}_d} \tag{23}$$

which has a Student distribution with $2n - 3$ degrees of freedom.

Based on the dispersion summation law, from equation (22) one obtains

$$s_{\bar{y}_d} = [(s^2)_{\bar{y}_0} + k^2(s^2)s_{\bar{y}}]^{\frac{1}{2}} \tag{24}$$

in which

$$(s^2)_{\bar{y}_0} = s_{\bar{y}}^2 \sum_{i=1}^{n} c_i^2 / n \sum_{i=1}^{n} (c_i - \bar{c})^2 \tag{25}$$

$$(s^2)_{s_{\bar{y}}} = s_{\bar{y}}^2 / 2(n - 2) \tag{26}$$

and

$$s_{\bar{y}}^2 = [\sum_{i=1}^{n} \bar{y}_i^2 - \bar{y}_0' \sum_{i=1}^{n} \bar{y}_i - b \sum_{i=1}^{n} \bar{y}_i c_i]/(n - 2) \tag{27}$$

represents the dispersion of values $\bar{y}_i$ around the calibration line (20).

Based on these considerations, the expression of the confidence interval for the analytical signal corresponding to the detection limit will be

$$\bar{y}_d' - t_{(P, 2n-3)} s_{\bar{y}_d} < \bar{y}_d < \bar{y}_d' + t_{(P, 2n-3)} s_{\bar{y}_d} \tag{28}$$

The confidence interval for the detection limit results from the inter-section of the confidence interval of the analytical signal corresponding to

543

the detection limit, with the confidence interval of the calibration function (20), viz. with the two hyperbolae (21). We shall give below an example of the manner in which the detection limit and its confidence interval are estimated from a series of calibration data.

Analytical signals of seven steel standard samples with a known chromium content have been recorded 31 times each, on an ARL-31 000 type emission automatic spectrometer. Working conditions: $\lambda_{Cr} = 2989.2\,\text{Å}$; argon; intermittent arc; recording time 3 s. The results of measurements (in digits) are listed in *Table 1*.

Table 1. Values of the analytical signal (digits) of seven steel standard samples, on ARL-31 000 type emission automatic spectrometer

Digits	Signal frequencies (number of samples)						
12	1	—	—	—	—	—	—
13	2	1	1	—	—	—	—
14	4	3	1	—	—	—	—
15	7	5	3	1	2	—	—
16	9	10	7	3	1	—	—
17	5	8	9	9	6	2	—
18	2	3	5	9	8	3	2
19	1	1	3	4	6	4	2
20	—	—	1	3	6	6	4
21	—	—	1	2	2	9	6
22	—	—	—	—	—	4	6
23	—	—	—	—	—	3	5
24	—	—	—	—	—	—	3
25	—	—	—	—	—	—	2
26	—	—	—	—	—	—	1
27	—	—	—	—	—	—	—
Digits ($\bar{y}$)	15.6	16.1	16.9	17.9	18.3	20.3	21.8
%Cr(x)	0.016	0.027	0.043	0.064	0.074	0.102	0.130

By calculation one finds: $\bar{y}_0' = 14.7$; $b = 53.4$; $s_{\bar{y}} = 1.75$; $(s^2)_{\bar{y}_0} = 0.0625$; $(s^2)_{s_{\bar{y}}} = 0.0080$; and $s_{\bar{y}_d} = 0.43$.

Therefore the estimate of the calibration function will be

$$\bar{y} = 14.7 + 53.4c \qquad (29)$$

Considering a false identification probability $P_{10} = 0.025$ ($z_k = 1.96$) as well as proper identification probability $P_{11} = 0.975$ ($z_d = 1.96$), i.e. a signal-to-noise ratio $k = 3.92$ (see equations (15) and (15')), one obtains $\bar{y}_d = 21.56$ (22) and therefore $\bar{c}_d' = 0.128$ per cent Cr (equation 29). Under the above-mentioned conditions, one finds $s_{\bar{y}_d} = 0.43$ (equation 24), so that for a confidence probability of 0.90, $t(0.90;431) = 1.65$, and thus, according to equation (28) one obtains

$$20.85 < \bar{y}_d < 22.27$$

From these two extreme values of the analytical signal, by introducing in
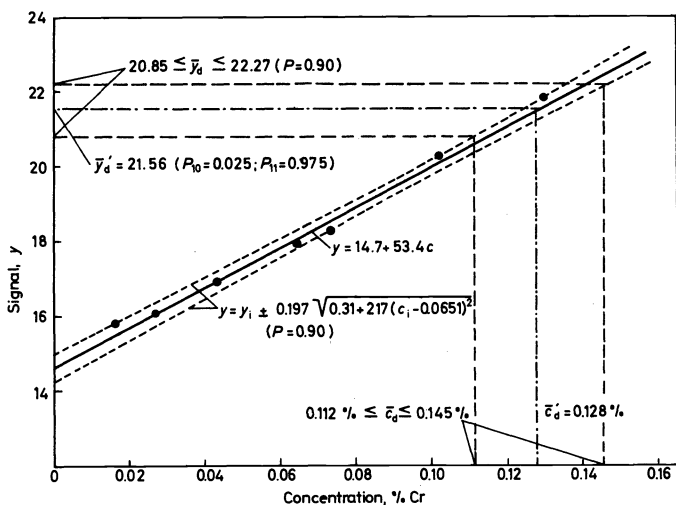
544

Figure 8. Evaluation of the detection limit for the experiment illustrated in *Table 1*.

equation (21) of the two hyperbolae, one finds for the detection limit the following confidence interval (see *Figure 8*):

$$0.112\% < \bar{c}_d < 0.145\%$$

From the above it results that the two steps of the model for estimating the detection limit (the probabilities P_{10} and P_{11}) are included in the coefficient $k = z_k + z_d$ (signal-to-noise ratio) of equation (22).

Estimation of the detection limit by the frequentometrical method

As has been shown, the method is based on the calculation of the concentration value for which the proper detection probability P_{11} reaches a certain value. As to the discrimination between the two alternatives, the component present or absent, there are two working procedures.

The first procedure[10] is based on a successive comparison (N repetitions) of the value of the analytical signal with that of the background (blind test) on n samples of known concentrations, but below the detection limit (uncertain reaction domain[25]). The second procedure[26] uses the two-step model and is based on the comparison of the values of the analytical signal of N repetitions with the value corresponding to the detection level y_k, calculated from the background fluctuations (blind tests):

$$y_k = \bar{y}'_0 + z_k s_{\bar{y}} \qquad (30)$$

considering a false identification probability P_{10} (error of the first kind). Based on the data given in *Table 1* and considering the same value for

$P_{10} = 0.025$ ($z_k = 1.96$) as well as equation (30), one obtains $y_k = 14.7 + 1.96 \times 1.75 = 18.2 \simeq 18$ digits.

In conclusion, all samples for which $y > 18$ will be positive and thus the primary analytical signal will be precisely the frequency of appearance of positive samples $F(c) = N_+/N$.

In view of linearization, one passes from frequencies to the variable z by means of the Laplace function $\Phi(z) = F(c) - 0.5$ for $F(c) > 0.5$ (positive z) and, correspondingly, $\Phi(z) = 0.5 - F(c)$ (negative z). Under these new conditions, the values of z will obviously play the role of the analytical signal. Since, according to Glivenko's theorem[27], the experimental frequencies obtained in this manner estimate the theoretical frequencies, the values z have obviously also an estimative character. Consequently a confidence interval must also be attached to function $z = f(c)$, i.e. a dispersion band limited by two hyperbolae described by equation (21). Based on the data of *Table 1*, one can derive *Table 2* and correspondingly *Figure 9*.

Table 2. Values of empirical frequencies and of variable z obtained from data of *Table 1* for a detection level $y_k = 18$ digits

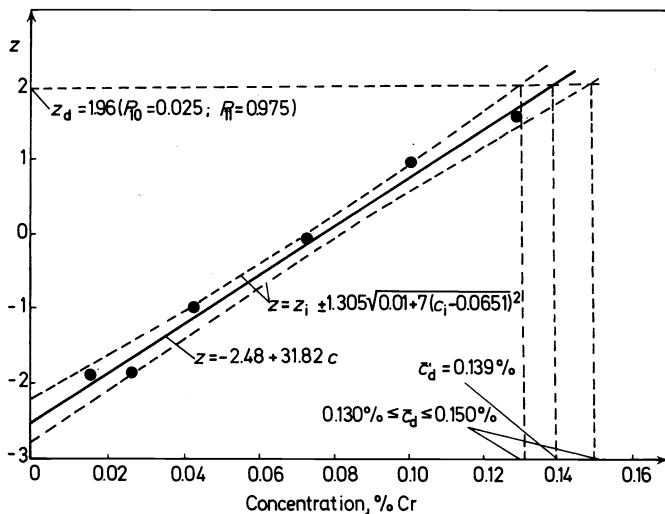| %Cr (x)	0.016	0.027	0.043	0.064	0.074	0.102	0.130
$F(c)$	0.032	0.032	0.160	0.290	0.450	0.840	0.935
$z(y)$	−1.85	−1.85	−1.00	−0.56	−0.12	+0.97	+1.53

Figure 9. Frequentometrical procedure for estimating the detection limit

From the data of *Table 2*, applying the least squares method, one obtains

$$z = 2.48 + 31.82c \qquad (31)$$

Considering, for example, $P_{11} = 0.975$ and thus $z_d = 1.96$, there results

for the detection limit, $\bar{c}'_d = 0.139$ per cent Cr, therefore a very close value to that obtained by the previously described method (0.128 per cent).

In order to calculate the confidence interval of the detection limit one proceeds as in *Figure 9*, since the value z_d results by univocal definition of the value of the proper identification probability P_{11}. For this reason the confidence interval of the detection limit results only from the uncertainty of the experimental function $z = f(c)$. Therefore, the two extreme values of the detection limit, as seen in *Figure 9*, are obtained by the intersection of value z_d with the two hyperbolae (21) associated with the calibration function, to give the expression

$$z = (-2.48 + 31.82c_i) \pm 2.02 \times 0.646[0.01 + 7(c_i - 0.0651)^2]^{\frac{1}{2}}$$

For $z = 1.96$ one obtains

$$0.130\% < \bar{c}_d < 0.150\%$$

There it results that the described frequentometrical procedure also includes two steps: the first, that the probability P_{10} is included in equation (30); and the second, that the probability P_{11} is included in the value z_d of the calibration function (31).

3. FORMULATION OF DETECTION DECISION BY MEANS OF STATISTICAL TESTS

Due to the random character of the analytical signal, the formulation of any detection decision must be compulsorily made by means of statistical criteria. We shall describe below the manner in which three tests should be applied with a view to checking the hypothesis concerning the detection of a component in the sample under investigation: a parametrical test (Student), a non-parametrical one (Wilcoxon) and a sequential test (Wald).

t Test (Student)

The '*t*' test may be used in formulating detection decisions by comparing a selection of analytical signals obtained on samples of the material under investigation with a selection of measurements obtained on samples free of the component to be detected, i.e. on blind samples (background).

To do this, one calculates the variable

$$t_{exp} = \frac{\bar{y}_i - \bar{y}_0}{\left\{ \frac{n_0 + n_i}{n_0 n_i} \left[\frac{(n_0 - 1)s_0^2 + (n_1 - 1)s_i^2}{n_0 + n_i - 2} \right] \right\}^{\frac{1}{2}}}$$

where s_0^2 and n_0, respectively, represent the dispersion and the number of measurements with mean $\bar{y}_0$ performed on the blind (background) sample and s_i^2 and n_i, respectively, represent the dispersion and the number of measurements with a mean $\bar{y}_i$ performed on the sample under investigation.

If $t_{exp} > t_{(P, n_0 + n_i - 2)}$[28], one accepts the hypothesis that the component to be detected is present, or, in general, that the material under investigation has a higher concentration on the component to be detected than the reference material.

Table 3 lists the results of spectral measurements performed on three steel

samples, a reference and two unknowns, to detect titanium. The results of the measurements represent the differences in blackening of segments on the spectral plate where one of the titanium lines must be present ($\lambda_{Ti} = 3103.8$ Å) together with that of the internal standard line ($\lambda_{Fe} = 3102.87$ Å).

Table 3. Spectral analytical measurements for titanium detection

No. (n)	Sample			Results
	black $\Delta S_0 \times 10^3$	for testing (1) $\Delta S_1 \times 10^3$	for testing (2) $\Delta S_2 \times 10^3$	
1	−265	−290	−240	
2	−262	−227	−236	
3	−254	−252	−232	$\bar{y}_0 = 258.5$
4	−266	−252	−253	$\bar{y}_1 = 255.3$
5	−225	−272	−222	$\bar{y}_2 = 238.4$
6	−246	−251	−243	$s_0^2 = 422.1$
7	−298	−227	−210	$s_1^2 = 419.0$
8	−251	−239	−241	$s_2^2 = 250.5$
9	−248	−250	−243	$t_{1,\,exp} = 0.37$
10	−287	−280	−271	$t_{2\,exp} = 2.58$
11	−292	−268	−231	

From Table 3 with the values of the variable t one finds $t_{(0.95,\,20)} = 2.09$ so that since $t_{l,\,exp} < t_{(0.95,\,20)}$ and $t_{2.\,exp} > t_{(0.95,\,20)}$ the hypothesis of the presence of titanium may be accepted only for sample 2.

Wilcoxon test[29]

In order to apply this test belonging to order statistics, a domain of mathematical statistics, the n_0 and n_i measurements performed, respectively, on the reference material and on the material under investigation are arranged according to the magnitude, obtaining a sequence of the form

$$y_0 y_0 y_i y_0 y_i y_0 y_0 y_i y_0 y_i \qquad (33)$$

If in such a sequence y_i appears after y_0, we shall say that one has an inversion. Thus sequence (33) contains 16 inversions.

One can show[29] that the mean M_u and the dispersion σ_u^2 of the number u of inversions, for the case in which the two series of measurements n_0 and n_i belong to the same distribution, are, respectively,

$$M_u = \tfrac{1}{2}n_0 n_i \qquad (34)$$

and

$$\sigma_u^2 = \tfrac{1}{12}n_0 n_i(n_0 + n_i + 1) \qquad (35)$$

For $(n_0, n_i) \geqslant 4$ and $(n_0 + n_i) \geqslant 20$, the distribution of the variable

$$z = (u - M_u)/\sigma_u \qquad (36)$$

may be considered as normal. If $z_{exp} \geqslant z_P$ (tabulated) one accepts the hypothesis that the component to be detected is present, or, in general, that

the concentration of the component to be detected has a higher value in the material under investigation than in the reference sample.

In order to apply the Wilcoxon test to the data of *Table 3*, the results of measurements for sample 1 under investigation are arranged in the order of their magnitude, forming the sequence

$$y_0 y_1 y_0 y_1 y_1 y_1 y_0 y_0 y_0 y_0 y_1 y_1 y_0 y_1 y_1 y_0 y_0 y_0 y_1 y_1 y_1 y_0$$

with $u_1 = 63$ inversions.

For sample 2 one obtains the sequence:

$$y_0 y_0 y_2 y_0 y_0 y_0 y_0 y_2 y_0 y_0 y_0 y_2 y_2 y_0 y_2 y_2 y_2 y_2 y_2 y_0 y_2 y_2$$

with $u_2 = 98$ inversions.

For $n_0 = n_1 = n_2$ in agreement with formulae (34) and (35) there results $M_u = 60$ and $\sigma_u^2 = 232$, and based on formula (36) one obtains $z_{1,\exp} = 0.13$ and $z_{2,\exp} = 2.5$. From a table with the distribution values of variable z^{23} one finds $z_{0.95} = 1.96$, so that since $z_{1,\exp} < z_{0.95}$ and $z_{2,\exp} > z_{0.95}$ the hypothesis of the presence of titanium may be accepted only for sample 2.

Sequential probability ratio test (Wald)[30]

The basic feature of sequential tests consists in the fact that the number of necessary measurements for decision-making depends on the result of observations itself. In other words, the very number (volume) of measurements is considered as a random variable.

Considering the two hypotheses H_1 with P_{11} and H_0 with P_{00} and, respectively, the two false detection probabilities P_{10} (error of the first kind) and P_{01} (error of the second kind), one first calculates the ratios

$$A = \frac{P_{11}}{P_{10}} = \frac{1 - P_{01}}{P_{10}} \text{ and } B = \frac{P_{01}}{P_{00}} = \frac{P_{01}}{1 - P_{10}} \tag{37}$$

Considering these ratios, the interval in which $\sum_{i=1}^{n} y_i$ will be found without being able to make a decision will be given by relation

$$\frac{\sigma_{\bar{y}}^2 \ln B}{\bar{y}_i - \bar{y}_0} + n \frac{\bar{y}_i + \bar{y}_0}{2} < \sum_{i=1}^{n} \bar{y}_i < \frac{\sigma_{\bar{y}}^2 \ln A}{\bar{y}_i - \bar{y}_0} + n \frac{\bar{y}_i + \bar{y}_0}{2} \tag{38}$$

where $\sigma_{\bar{y}}^2$ is the dispersion of the analytical signal around the calibration line (20) (it is calculated by formula 26), $\bar{y}_0$ is the intercept of the calibration function (20), and $\bar{y}_i$ is the value of the analytical signal (calculated from function 20) for the reference concentration (subjected to decision).

Therefore, in the case of a practical detection problem framed in such a sequential model, based on the result of the repeated measurement process, one calculates the consecutive sums of measurement results, comparing them with the two decision levels in agreement with relation (38):

(1) If $\sum_{i=1}^{n} \bar{y}_i \leqslant \frac{\sigma_{\bar{y}}^2 \ln B}{\bar{y}_i - \bar{y}_0} + n \frac{\bar{y}_i + \bar{y}_0}{2}$ $\qquad$. $\tag{39}$

one accepts hypothesis H_0.

$$(2) \text{ If } \sum_{i=1}^{n} \bar{y}_i \geqslant \frac{\sigma_{\bar{y}}^2 \ln A}{\bar{y}_i - \bar{y}_0} + n\frac{\bar{y}_i + \bar{y}_0}{2} \qquad (40)$$

one accepts hypothesis H_1.

(3) If neither hypothesis (39) nor (40) is satisfied, the decision will be made to continue the experiment.

Further, based on the data listed in *Table 4*, we give an example of how to apply the test for the detection of tungsten in steel.

Table 4. Calibration data for tungsten (λ_w = 2079.1Å) for five steel samples obtained with an ARL-31 000 type automatic emission spectrometer

Sample No.

1	2	3	4	5
		Concentration (%W)		
x = 0.053	0.022	0.012	0.006	0.001
222	79	36	20	10
221	75	53	12	5
215	68	42	4	3
223	74	53	24	8
205	71	59	19	3
203	83	42	24	13
227	68	50	19	5
197	66	59	25	0
211	91	57	15	1
196	71	46	20	16
216	90	60	18	12
226	103	55	17	2
222	80	52	29	5
$\bar{y}$ = 214.2	78.4	51.1	18.9	6.4

By calculation one finds $\bar{y}_0 = -1.8$ and $b = 4025.9$, so that the calibration function will have the form

$$\bar{y} = -1.8 + 4025.9c \qquad (41)$$

With $\sigma_y^2 = 94.7$ (since $n = 5 \times 13 = 65$) one considers that the dispersion does characterize the general collectivity.

Under similar analytical conditions to those in which the calibration data of *Table 4* were obtained, a steel sample was subjected to a repeated measuring process with a view to establishing the significance of its tungsten content, in relation to hypotheses: the sample does not contain tungsten (H_0) and the concentration of tungsten is equal to or higher than 0.002 per cent (the sample contains tungsten). From the calibration function (41) obtained from the data of *Table 4*, for $c_w = 0.002$ per cent, there results $\bar{y}_i = 6.25$ digits. Considering $P_{10} = 0.025$ ($P_{00} = 0.975$), $P_{01} = 0.05$ ($P_{11} = 0.95$), there results $A = 38$, $B = 0.0513$. In conclusion, in agreement with equation (39), if

$$\sum_{i=1}^{n} \bar{y}_i \leqslant -34.9 + 2.23\,n \qquad (42)$$

550

Table 5. Formulating of decisions by means of the sequential probability ratio test for a black sample of steel in relation to hypotheses H_0 ($c < 0.002$ per cent W) and H_1 ($c \geqslant 0.002$ per cent)

No. of measurements (n)	Result of measurements	$\sum\limits_{i=1}^{n} y_i$	Decision level for H_0 hypothesis	Decision level for H_1 hypothesis
0	0	0	−34.9	42.7
1	8	8	−32.7	44.9
2	6	14	−30.4	47.2
3	9	23	−28.2	49.4
4	5	28	−26.0	51.6
5	9	37	−23.8	53.9
6	8	45	−21.5	56.1
7	7	52	−19.3	58.3
8	9	61	−17.1	60.5
9	6	67	−14.8	62.8
10	2	69	−12.6	65.0
11	13	82	−10.4	67.2

one accepts hypothesis H_0 ($c_W \leqslant 0.002$ per cent) and, in agreement with (40), if

$$\sum_{i=1}^{n} \bar{y}_i \geqslant 42.7 + 2.23\,n \tag{43}$$

one accepts hypothesis H_1 ($c_W > 0.002$ per cent). If

$$-34.9 + 2.23\,n < \sum_{i=1}^{n} \bar{y}_i < 42.7 + 2.23\,n \tag{44}$$

the experiment will continue (indifference zone).

Table 5 illustrates that beginning with the eighth measurement, inequality (43) is fulfilled and therefore hypothesis H_1 must be accepted, i.e. $c_W > 0.002$ per cent.

The manner in which detection decisions could be formulated by means of the sequential probability ratio test is shown in a graphical form in *Figure 10*.

It follows from *Figure 10* that, starting with the eighth measurement, the sums of values of consecutive results leave the domain where one decides on the continuation of the experiment and enter the domain in which one decides on acceptance of hypothesis H_1 ($c_W \geqslant 0.002$ per cent). A correct estimation of the detection limit is based on a mechanism presented in *Figure 11*.

Since, as has already been shown (equation 1), each measurement depends on a very large number of variables, the analytical signal, like the background signal, in fact, will always have a random character with a distribution described by a certain limit theorem. In conclusion, as a result of the action of perturbations, the passage from the space of the sample under investigation (space of hypotheses) to the space of measurement results (process 1) takes place by a mechanism governed by probabilistic laws. Under such conditions, the passage from the space of measurement results to the decisions space (process 2) must necessarily involve certain decision rules based on statistical criteria, as shown previously.
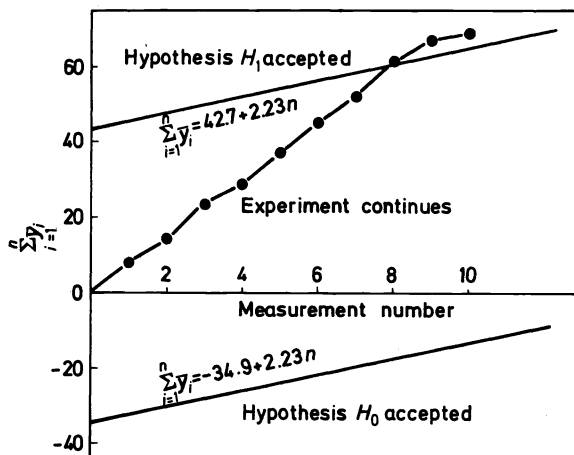
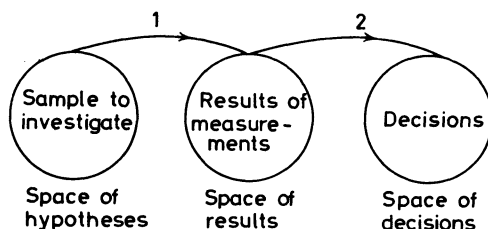*Figure 10.* Detection by means of the sequential probability ratio test

Figure 11. Mechanism of the analytical detection process

The statistical theory of the detection of signals, largely developed in recent times because of its scientific implications (spatial, in particular), in communication, in transportation and also in other domains, actually constitutes the base of the definition for estimating and making decisions in the field regarding both detection limit and determination limit.

As shown before, the result of the estimation of the detection limit, and in fact the expression of any results, depends on the accepted probabilities. A standardization is therefore necessary in this domain and it could be organized by the IUPAC.

In conclusion, taking into account the statistical theory of detection limit with all its implications and also the tremendous importance of trace analysis in the most varied fields of science, technology and biology, it can be stated that the problem as a whole should be reconsidered by the IUPAC Analytical Chemistry Division.

REFERENCES

[1] H. Kaiser. *Spectrochim. Acta.* **3.** 40 (1947).
[2] H. Kaiser and H. Specker. *Z. Anal. Chem.* **149.** 46 (1956).

3 G. Gottschalk and P. Dehmel, *Z. Anal. Chem.* **160**, 167 (1958).
4 V. V. Nalimov and V. V. Nedler, *Zavodsk. Lab.* **27**, 861 (1961).
5 A. L. Wilson, *Analyst,* **86**, 72 (1961).
6 J. B. Ross, *Analyst,* **87**, 832 (1962).
7 H. Kaiser, *Z. Anal. Chem.* **209**, 1 (1965).
8 H. Kaiser, *Z. Anal. Chem.* **216**, 80 (1966).
9 G. Ehrlich and H. Mai, *Z. Anal. Chem.* **218**, 1 (1966).
10 C. Liteanu and I. Florea, *Mikrochim. Acta,* 983 (1966).
11 H. Specker, *Z. Anal. Chem.* **221**, 33 (1966).
12 D. J. Hobbs and D. M. Smith, *Canad. Spectroscopy,* **11**, 5 (1966).
13 K. Doerffel, *2nd Int. Symposium, Reinstoffe Wiss. Techn.,* Dresden (1965), Vol. 2, p 19 (1966).
14 P. A. St. John, W. J. McCartty and C. D. Winefordner, *Anal. Chem.* **39**, 1945 (1967).
15 G. Erhlich, *Z. Anal. Chem.* **232**, 1 (1968).
16 V. Svoboda and R. Gerbatsch, *Z. Anal. Chem.* **242**, 1 (1968).
17 R. Püschel, *Mikrochim. Acta,* 82 (1968).
18 L. A. Currie, *Anal. Chem.* **40**, 586 (1968).
19 D. J. Hobbs and A. Iny, *Appl. Spectrosc.* **24**, 522 (1970).
20 C. Liteanu and I. Rîcă, *Mikrochim. Acta,* 745 (1973).
21 C. W. Helstrom, *Statistical Theory of Signal Detection,* p 59. Pergamon: Oxford (1960).
22 W. L. Root, *Proc. IEEE,* **58**, 610 (1970).
23 P. D. Lark, B. R. Craven and R. C. L. Bosworth, *The Handling of Chemical Data,* p 335. Pergamon: Oxford (1969).
24 F. S. Acton, *Analysis of Straight-line Data,* p 34. Wiley: New York (1959).
25 F. Emich, *Ber. Dtsch. Chem. Ges.* **43**, 10 (1910).
26 C. Liteanu and I. Rîcă, *Mikrochim. Acta,* 311 (1975).
27 B. V. Gnedenko, J. K. Beliaev and A. D. Soloviev, *Matematicheskie Metody v Teorii Nadeshnosti,* p 59. Izd. Nauka: Moscow (1965).
28 J. Mandel, *The Statistical Analysis of Experimental Data,* p 393. Interscience: New York (1964).
29 I. Vincze, *Mathematische Statistik mit industriellen Anwendungen,* p 179. Akad. Kiadó: Budapest (1971).
30 C. W. Helstrom, *Statistical Theory of Signal Detection,* p 79. Pergamon: Oxford (1960).