That (7) is indeed satisfied now follows from (5) and the finiteness of the function $l(t)$ since for a large enough integer M one has

$$\sum_{N=M}^{\infty} P(n = N \mid F) \exp [r_1 \log N - N \log \varphi(t_0)]$$

$$\leqq \sum_{N=M}^{\infty} P(n = N \mid F) \exp [Nt_1 - N \log \varphi(t_0)] < \infty.$$

Thus the expected value on the extreme right in (8) is finite. This completes the proof of the theorem.

REFERENCES

[1] A. WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283–285.

[2] A. WALD, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 493–496.

[3] CHARLES STEIN, "A note on cumulative sums," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.

A SIGNIFICANCE TEST AND ESTIMATION IN THE CASE OF EXPONENTIAL REGRESSION

BY D. S. VILLARS[1]

United States Rubber Company, Passaic, N. J.

1. Introduction. The principal problem under consideration in this note may be described as follows. Consider a variate, z, whose distribution for a given value of a fixed variate, t, is:

$$f(z \mid t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(z-a+be^{-kt})^2/2\sigma^2} \tag{1.1}$$

where a, b, and k are real-valued parameters. The regression of z on t is exponential, for it follows from (1.1) that the expected value of z, given t, is:

$$E(z \mid t) = a - be^{-kt}. \tag{1.2}$$

On the basis of a random sample $0_N(z_1, t_1 ; z_2, t_2 ; \cdots ; z_N, t_N)$ it is desired to test whether $k = 0$ or ∞. The problem of "fitting" a curve, $z = a - be^{-kt}$, to the sample (*i. e.* of estimating a, b, and k from the sample) will also be treated.

As an illustration of how the statistical problems described above arise in

[1] Present address, Jersey City Junior College, Jersey City, N. J.

practice, let us consider a typical situation in industrial chemistry. Let the quantity, z, be a property of a latex and let the quantity, t, be time. Suppose, furthermore, that measurement of t is without error but that measurement of z is subject to error; let it be assumed that the observed value in a measurement of z is a variate having a normal (Gaussian) distribution about the "true value," $E(z)$. On basis of N independent measurements, $z_1, z_2, \cdots, z_N$ of z at times, $t_1, t_2, \cdots, t_N$, respectively, the experimenter may wish to test the hypothesis that $k = 0$ or ∞. If this hypothesis is true the suspected exponential relation between z and t does not hold; in this case $E(z)$ is a constant ($a - b$, or a) and estimation of the constant from the data is quite straightforward. If the data conflict with the hypothesis that $k = 0$ or ∞, the experimenter may wish to estimate the parameters, a, b, and k (*i. e.*, "fit" the curve, $z = a - be^{-kt}$, to the data).

The problems considered in this note will be treated only for the case where N is an even integer (≥ 6) and the times $t_1, t_2, \cdots, t_N$ at which measurements of z are made are such that

$$t_{2\alpha} - t_{2\alpha-1} = \Delta, \qquad \text{a constant, } (\alpha = 1, 2, \cdots, n = N/2). \tag{1.3}$$

The odd time intervals, $t_3 - t_2, t_5 - t_4$, etc. do not have to be equal.

2. Test of the hypothesis that $k = 0$ or ∞. The space, say Ω, of admissible values of the parameters in (1.1) is: $\sigma^2 > 0, -\infty < a, b, k < +\infty$. Under the null hypothesis the admissible values of the parameters lie in a subspace of Ω, say ω, specified as follows: $\sigma^2 > 0, -\infty < a, b < +\infty, k = 0$, or ∞.

Let $y_j = z_{2\alpha}$ and $x_j = z_{2\alpha-1}$, $(\alpha = 1, \cdots, n = N/2)$. From (1.1) and (1.3) it follows that the n pairs x_j, y_j are normally and independently distributed with common variance, σ^2, that x_j and y_j are independent ($j = 1, 2, \cdots, n$), and that

$$\nu_j = h + m\mu_j \tag{2.1}$$

where $\nu_j = E(y_j)$, $\mu_j = E(x_j)$, $h = a(1 - e^{-k\Delta})$, and $m = e^{-k\Delta}$. The space, Ω', of admissible values of the parameters in the joint distribution of x_j, y_j, $(j = 1, \cdots, n)$, is: $\sigma^2 > 0, \nu_j = h + m\mu_j, -\infty < h < +\infty, -\infty < \mu_j, \nu_j < +\infty; 0 \leq m < \infty$. The subspace of Ω', say ω', associated with the null hypothesis is: $\sigma^2 > 0, \nu_j = \mu_j = c$, where $c = a - b$ or a according as $k = 0$ or ∞. In Ω', the expected values of x and y lie on a line; in ω' they lie in a single point. It is clear that by transforming the original sample $0_N(z_1, t_1, \cdots, z_N, t_N)$ to a sample $0_n(x_1, y_1; \cdots; x_n, y_n)$ we have reduced the original problem to the familiar problem of linear regression in which there is "error in both variates".

The slope of the "line of best fit" to the sample points $(x_1, y_1; \cdots; x_n, y_n)$ is [1]:

$$\hat{m} = [S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}]/2S_{xy} \tag{2.2}$$

where

$$S_{xx} \equiv \sum_1^n (x_j - \bar{x})^2$$

$$S_{xy} \equiv \sum_1^r (x_j - \bar{x})(y_j - \bar{y})$$

$$S_{yy} \equiv \sum_1^n (y_j - \bar{y})^2$$

$$\bar{x} \equiv \sum_1^n x_j/n$$

$$\bar{y} \equiv \sum_1^n y_j/n$$

($\hat{m}$ is an estimate of m in (2.1)). Since $m = e^{-k\Delta}$ (where k and Δ are real), it is intuitively clear that when m is non-positive the sample 0_n does not conflict with the null hypothesis. The null hypothesis can be tested by means of the statistic [2, 144]

$$F' = \frac{S_{xx} + 2mS_{xy} + m^2 S_{yy}}{m^2 S_{xx} - 2mS_{xy} + S_{yy}}. \tag{2.3}$$

The null hypothesis is rejected if $\hat{m}$ is positive and F' is large. Percentage points of the distribution of F' are given in [2, 146] for $n = 3$ (1) 15 (5) 30, 40, 60, 120 and for significance levels, 0.001, .01, .05, .10, and .20. These significance levels, however, were computed for use in cases where the sign of $\hat{m}$ was irrelevant. It happens that to test the null hypothesis under consideration in this problem at a significance level α we should use a critical value of F' (given in [2]) corresponding to a significance level 2α. The reason for this is that when the null hypothesis is true the quantities m and F' are independent and the probability that $\hat{m}$ is positive is $\frac{1}{2}$—thus the chance of rejecting the null hypothesis is $\frac{1}{2}(2\alpha) = \alpha$.

3. Estimation of a, b, and k. If the data do not support the hypothesis that $k = 0$ or ∞, the experimenter may wish to estimate a, b, and k. General alternative methods of estimating these parameters will now be considered.

(1) Estimate a, b, and k from 0_N by the method of least squares; *i.e.*, solve the simultaneous equations $\partial S/\partial a = 0$, $\partial S/\partial b = 0$, and $\partial S/\partial k = 0$ for a, b, and k, where

$$S = \sum_{i=1}^N (z_i - a + be^{-kt_i})^2. \tag{3.1}$$

The value of k obtained by this method of estimation will not in general be the same as that computable from $\hat{m}$ in (2.2) and used for the significance testing.

(2) Estimate k by means of (2.2) and the relation $m = e^{-k\Delta}$, then substitute this estimate into S of (3.1) and estimate a and b by means of least squares.

(3) Estimate k as in (2) and choose, as an estimate of a, the intercept of the "line of best fit" for 0_n. Then substitute these estimates of a and k into (3.1) and estimate b by means of least squares. In this case the estimate of b comes out to be:

$$\hat{b} = \sum_1^N e^{-\hat{k}t_i}(\hat{a} - z_i) \Big/ \sum_1^N e^{-2\hat{k}t_i} \tag{3.2}$$

where $\hat{a}$ and $\hat{k}$ are the estimates of a and k.

If the values, $t_1, t_2, \cdots, t_N$ are such that $t_{i+1} - t_i = \Delta$, $(i = 1, 2, \cdots, N-1)$, the following estimation procedure might be used.

(4) Let

$$y_j = z_{i+1}$$
$$x_j = z_i \qquad (i = 1, 2, \cdots, N-1),$$

and treat the $(N-1)$ pairs of values $(x_1, y_1; \cdots; x_{N-1}, y_{N-1})$ as a sample of size $(N-1)$. Using this sample, estimate k, a, and b in a manner similar to that in (2) or (3). It should be noted that this sample is not a random sample owing to the dependence among the $(N-1)$ elements.

The procedure in alternative (1) is very laborious and time-consuming. The procedure in (2) and (3) can be carried out quickly and easily. In (1) the method of least squares yields the same results as would be obtained from application of the method of maximum likelihood. Examples of estimation by procedures (3) and (4) are given in the next section.

4. Example. The accompanying table lists experimentally observed values of a property of a latex obtained at biweekly intervals. Using the first, third, *etc.*, quantities as x_j and the remaining ones as y_j, the sums of squares and products of deviations are found to be:

$$S_{xx} = .035510 \qquad \bar{x} = 0.9195$$
$$S_{xy} = .025645$$
$$S_{yy} = .023414 \qquad \bar{y} = .9365.$$

Substituting these values in equation (2.2) and computing the other constants from equation (2.1) we get: $m = 0.791596$, $a = 1.0009$, and $k = 0.1168$. The F' ratio is (2.3) 17.03. Entering Table I of [2], we find that for eight point pairs a value of $F' = 16.5$ may be expected only one time in one hundred. On excluding the possibility of negative values of m, this corresponds to the 0.5% significance level. The exponential relationship is thus concluded to be highly significant.

Evaluation of b by equation (3.2), method 3, gives 0.2560, if all 16 values are used. The equation calculated from the data is thus:

$$z = 1.0009 - 0.2560\, e^{-0.1168t}. \tag{4.1}$$

The alternative procedure, method 4, would be to use all the z_i points for the estimation of a and k. This leads to the following values of the computation quantities:

$$S_{xx} = \sum_{i=1}^{16} x_i^2 - x_{16}^2 = 0.052374; \qquad \bar{x} = 0.9223$$

$$S_{xy} = \sum_{i=1}^{15} x_i x_{i+1} = .036924$$

$$S_{yy} = \sum_{i=1}^{16} x_i^2 - x_1^2 = .035436; \qquad \bar{y} = .9381.$$

Note that the difference $S_{yy} - S_{xx}$ used in the formula for m cancels out all intervening squares between the first and last.

$$S_{yy} - S_{xx} = x_1^2 - x_{16}^2.$$

TABLE I

t weeks	z_i	t weeks	z_i	t weeks	z_i	t weeks	z_i
1	.776	9	.939	17	.942	25	.955
3	.852	11	.904	19	.938	27	.993
5	.850	13	.930	21	.979	29	.985
7	.869	15	.948	23	.975	31	1.013

However, the data excluded thereby are in effect included in the new S_{xy}.

The final values obtained by the fourth procedure are: $m = 0.796596$, $a = 1.0000$, and $k = 0.1137$. The writer does not know whether the peculiar transference of data from $S_{yy} - S_{xx}$ to S_{xy} characteristic of procedure 4 improves the accuracy of the fit or hurts it. It is his personal preference to use procedure 3.

5. Acknowledgement. The writer wishes to acknowledge with thanks his gratitude to Drs. T. W. Anderson, Jr. and David F. Votaw, Jr. for many suggestions and discussions concerning this problem and for much help in clarifying the presentation of the concepts.

REFERENCES

[1] CHARLES H. KUMMELL, *The Analyst* (Des Moines), Vol. 6 (1879), pp. 97–105.

[2] D. S. VILLARS AND T. W. ANDERSON, JR., "Some significance tests for normal bivariate distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 141–148.