Normalization and statistical noise level in the normalized autocorrelation function. Compensated normalization

1991 J. Phys. A: Math. Gen. 24 L225

(http://iopscience.iop.org/0305-4470/24/5/003)

J. Phys. A: Math. Gen. 24 (1991) L225–L229. Printed in the UK

LETTER TO THE EDITOR

Normalization and statistical noise level in the normalized autocorrelation function. Compensated normalization*

Zbigniew Kojro†

Institute of Experimental Physics, University of Gdansk, Poland

Received 31 October 1990

Abstract. The use of large delay times in modern correlation experiments must be supported by an appropriate normalization scheme. The already available symmetric normalization does not eliminate all those statistical errors from a measured intensity autocorrelation function which can be eliminated with the help of a normalization procedure. A new kind of 'compensated normalization' is presented.

The symmetric normalization of the intensity autocorrelation function, suggested by Schätzel *et al* [1], diminishes statistical errors in a measured autocorrelation function in comparison with the standard, asymmetric normalization. There exists a different kind of normalization that leads to an even greater noise decrease. The formulation of the problem proposed by Schätzel *et al* [1] is generally valid (not only for large sample times) and allows us to see more details that are important in correlation measurements. Based on papers [1–3] one can write the following definitions and relations (notation used in these papers):

$$n_m = \bar{n}(1+\delta_m) \qquad \text{(photon counting signal at time } t_m \text{ [1])} \tag{1}$$

$$\delta B_\varepsilon = \frac{2}{N}\sum_{m=1}^{N} \delta_m \tag{2}$$

$$\varepsilon_k^{\text{lin}} = \frac{1}{N}\sum_{m=1}^{k} (\delta_{m+N} - \delta_m) \tag{3}$$

$$\varepsilon_k^{\text{sq}} = \frac{1}{N}\sum_{m=1}^{N} \delta_m\delta_{m+k} - \chi_k^2 \tag{4}$$

$$\chi_k^2 = \langle \delta_0 \delta_k \rangle \qquad \text{(square of expected field correlations [2])} \tag{5}$$

$$\varepsilon_{k,r} = \varepsilon_k^{\text{lin}} + \varepsilon_k^{\text{sq}} \qquad \text{(reduced error fluctuations [3])} \tag{6}$$

$$\varepsilon_k = \delta B_\varepsilon + \varepsilon_{k,r} \qquad \text{(total error fluctuations [3])} \tag{7}$$

$$\hat{G}_k^{(2)} = \frac{1}{N}\sum_{m=1}^{N} n_m n_{m+k} = \bar{n}^2(1+\chi_k^2+\varepsilon_k) \qquad \text{(see [3])} \tag{8}$$

$$\hat{n} = \frac{1}{N}\sum_{m=1}^{N} n_m = \bar{n}(1+\tfrac{1}{2}\delta B_\varepsilon) \qquad [1,2] \tag{9}$$

$$\hat{n}_k = \frac{1}{N}\sum_{m=1}^{N} n_{m+k} = \bar{n}(1+\tfrac{1}{2}\delta B_\varepsilon + \varepsilon_k^{\text{lin}}) \qquad [1] \tag{10}$$

* This work was supported within the project CPBP 01.06.2.01 (Poland).

† Address for correspondence: Max-Planck Institut für Biophysik, Kennedy Allee 70, 6000 Frankfurt/M 70, Federal Republic of Germany.

0305-4470/91/050225+05$03.50

$$\delta B = \frac{\hat{n}^2 - \bar{n}^2}{\bar{n}^2} = 2\frac{\hat{n} - \bar{n}}{\bar{n}} + \frac{1}{4}\left(2\frac{\hat{n} - \bar{n}}{\bar{n}}\right)^2$$

$$= \delta B_\varepsilon + \tfrac{1}{4}(\delta B_\varepsilon)^2 \approx \delta B_\varepsilon \qquad \text{(baseline error [3])} \tag{11}$$

$$\delta B_k = \frac{\hat{n}\hat{n}_k - \bar{n}^2}{\bar{n}^2}$$

$$= \delta B + \frac{\hat{n}(\hat{n}_k - \hat{n})}{\bar{n}^2} = \delta B_\varepsilon + \tfrac{1}{4}(\delta B_\varepsilon)^2 + \varepsilon_k^{\text{lin}}$$

$$+ \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}} \approx \delta B_\varepsilon + \varepsilon_k^{\text{lin}} \qquad \text{(symmetrized baseline error)} \tag{12}$$

where

$$\langle \delta B_\varepsilon \rangle = \langle \varepsilon_k^{\text{lin}} \rangle = \langle \varepsilon_k^{\text{sq}} \rangle = 0 \qquad \langle \delta B \rangle \approx 0 \qquad \langle \delta B_k \rangle \approx 0. \tag{13}$$

From (8)–(10) we obtain the unbiased, normalized estimator of intensity correlations [4]

$$\hat{g}_{0k} = \hat{G}_k^{(2)}/\bar{n}^2 = \{1 + \chi_k^2\} + \{\delta B_\varepsilon + \varepsilon_k^{\text{sq}} + \varepsilon_k^{\text{lin}}\} \tag{14}$$

and two biased estimators: the standard, asymmetrically normalized estimator [2, 5]

$$\hat{g}_k = \hat{G}_k^{(2)}/\hat{n}^2 = \frac{1 + \chi_k^2 + \delta B_\varepsilon + \varepsilon_k^{\text{sq}} + \varepsilon_k^{\text{lin}}}{1 + \delta B_\varepsilon + \frac{1}{4}(\delta B_\varepsilon)^2}$$

$$\approx \{1 + \chi_k^2\} + \{\varepsilon_k^{\text{sq}} + \varepsilon_k^{\text{lin}} - \delta B_\varepsilon \chi_k^2\}$$

$$+ \{\tfrac{3}{4}(\delta B_\varepsilon)^2 \chi_k^2 - \delta B_\varepsilon \varepsilon_k^{\text{sq}} - \delta B_\varepsilon \varepsilon_k^{\text{lin}} - \tfrac{1}{4}(\delta B_\varepsilon)^2\} \tag{15}$$

and the symmetrized one [1]

$$\hat{g}_k^{(s)} = \hat{G}_k^{(2)}/(\hat{n}\hat{n}_k)$$

$$= \frac{1 + \chi_k^2 + \delta B_\varepsilon + \varepsilon_k^{\text{sq}} + \varepsilon_k^{\text{lin}}}{1 + \delta B_\varepsilon + \varepsilon_k^{\text{lin}} + \frac{1}{4}(\delta B_\varepsilon)^2 + \frac{1}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}}}$$

$$\approx \{1 + \chi_k^2\} + \{\varepsilon_k^{\text{sq}} - \varepsilon_k^{\text{lin}} \chi_k^2 - \delta B_\varepsilon \chi_k^2\}$$

$$+ \{\tfrac{3}{4}(\delta B_\varepsilon)^2 \chi_k^2 - \delta B_\varepsilon \varepsilon_k^{\text{sq}} - \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}} - \tfrac{1}{4}(\delta B_\varepsilon)^2$$

$$+ \tfrac{3}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}} \chi_k^2 + (\varepsilon_k^{\text{lin}})^2 \chi_k^2 - \varepsilon_k^{\text{sq}} \varepsilon_k^{\text{lin}}\}. \tag{16}$$

Equation (15) is identical to (13) from [3] after using definition (11).

Taking into account the relation (equations (9) and (10))

$$\hat{n}_k - \hat{n} = \bar{n}\varepsilon_k^{\text{lin}} \tag{17}$$

one can introduce a new kind of 'compensated normalization' that gives us the third biased estimator

$$\hat{g}_k^{(c)} = [\hat{G}_k^{(2)} - (\hat{n}_k - \hat{n})\hat{n}]/\hat{n}^2$$

$$= \frac{1 + \chi_k^2 + \delta B_\varepsilon + \varepsilon_k^{\text{sq}} - \frac{1}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}}}{1 + \delta B_\varepsilon + \frac{1}{4}(\delta B_\varepsilon)^2}$$

$$\approx \{1 + \chi_k^2\} + \{\varepsilon_k^{\text{sq}} - \delta B_\varepsilon \chi_k^2\}$$

$$+ \{\tfrac{3}{4}(\delta B_\varepsilon)^2 \chi_k^2 - \delta B_\varepsilon \varepsilon_k^{\text{sq}} - \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\text{lin}} - \tfrac{1}{4}(\delta B_\varepsilon)^2\}. \tag{18}$$

Three kinds of terms are grouped in expressions (14), (15), (16) and (18) in { } brackets: expected values, statistical errors of first order of magnitude and statistical errors of second order of magnitude (bias terms), respectively (terms less than those of second order of magnitude are neglected). One can say that each normalization procedure removes a part of error contributions of first order of magnitude from an unbiased autocorrelation estimator (8) and simultaneously adds to it errors of second order of magnitude (bias terms). The quality of the normalization procedure can then be characterized by the degree of removing of undesirable terms from a measured autocorrelation function. We have from (15), (16) and (18):

(1) The standard, asymmetric normalization removes the baseline error [3].

(2) The symmetric normalization removes the baseline error and linear error contributions for delay times much larger than a coherence time. This has been predicted and experimentally checked by Schätzel *et al* [1] (although some terms of order $O(\delta^3)$, which contribute to the error of first order of magnitude and to the bias, have not been taken into account).

(3) The compensated normalization removes baseline error and linear error contributions in the whole delay time range.

(4) The term $-\delta B_\varepsilon \chi_k^2$ is insignificant in size distribution or linewidth determination [3].

Equations (15), (16) and (18) can be rewritten as

$$\hat{g}_k - 1 = (\hat{G}_k^{(2)} - \hat{n}^2)/\hat{n}^2 = \hat{G}_{k,\mathrm{M}}^{(2)}/\hat{n}^2$$
$$= A[\{\chi_k^2\} + \{\varepsilon_k^{\mathrm{sq}} + \varepsilon_k^{\mathrm{lin}}\} - \{\tfrac{1}{4}(\delta B_\varepsilon)^2\}] \tag{19}$$

$$\hat{g}_k^{(s)} - 1 = (\hat{G}_k^{(2)} - \hat{n}\hat{n}_k)/(\hat{n}\hat{n}_k) = \hat{G}_{k,\mathrm{c}}^{(2)}/(\hat{n}\hat{n}_k)$$
$$= A_k[\{\chi_k^2\} + \{\varepsilon_k^{\mathrm{sq}}\} - \{\tfrac{1}{4}(\delta B_\varepsilon)^2 + \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\mathrm{lin}}\}] \tag{20}$$

$$\hat{g}_k^{(c)} - 1 = (\hat{G}_k^{(2)} - \hat{n}\hat{n}_k)/\hat{n}^2 = \hat{G}_{k,\mathrm{c}}^{(2)}/\hat{n}^2$$
$$= A[\{\chi_k^2\} + \{\varepsilon_k^{\mathrm{sq}}\} - \{\tfrac{1}{4}(\delta B_\varepsilon)^2 + \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\mathrm{lin}}\}] \tag{21}$$

where

$$A = 1/(1 + \delta B_\varepsilon + \tfrac{1}{4}(\delta B_\varepsilon)^2) \tag{22}$$

$$A_k = 1/(1 + \delta B_\varepsilon + \varepsilon_k^{\mathrm{lin}} + \tfrac{1}{4}(\delta B_\varepsilon)^2 + \tfrac{1}{2}\delta B_\varepsilon \varepsilon_k^{\mathrm{lin}}). \tag{23}$$

Here $\hat{G}_{k,\mathrm{M}}^{(2)}$ is the so-called 'modified estimator' [6], and $\hat{G}_{k,\mathrm{c}}^{(2)}$ we call 'compensated estimator'.

Normalization is usually understood as a division by a number. Such an action gives us a singularity of amplitude, field under a curve, etc. In size distribution or linewidth determination, however, this procedure is not necessary [3] and the only important procedure is to subtract a background from a measured autocorrelation function. On the other hand, a division by $\hat{n}^2$ is very useful because it allows us to keep the correlations in particular correlator channels on the same level—however the choice of this normalization number is arbitrary. From (20) we see that in the case of symmetric normalization the denominator (and also factor A_k) is not a constant number and this causes the appearance of the term $-\varepsilon_k^{\mathrm{lin}}\chi_k^2$ in (16). This means that the error compensation resulting from the subtraction of $\hat{n}\hat{n}_k$ from $\hat{G}_k^{(2)}$ is partially lost in this case: the error is overcompensated. Comparing (19) and (21) with (15) and (18) we see that the only important errors are those from (19) and (21), and in size distribution or linewidth determination it is enough to treat the final factor A in (19) and (21) as an ordinary number which has no influence on the final result [3]. From (19) and (21) we also see that the normalization procedure for autocorrelation functions consists of the compensation and a typical normalization being the division by an ordinary number.

There exists the need for a detailed discussion of the contribution of error term $\varepsilon_k^{\text{lin}}$ *to the final statistical distortion of the measured autocorrelation function.* From (2) we see that when $k \ll N$ (where $k = \tau_k / T_s$, T_s is the sample time and τ_k the delay time) this error term is of order N^{-1} and therefore is much smaller than other error terms (2) and (4) which are of order $N^{-1/2}$ [2–5, 7]. This is a typical situation for correlators with linear delay time spacing, and this term does not contribute to variances and covariances calculated for statistical noise in autocorrelation function [2–5, 7]. At multiple sample and delay times [6, 8, 9] we have a quite different situation and this condition is no longer valid (in this case we have for the last delay time $\tau_{\max}$: $k = N$). A straightforward calculation gives us the standard deviation for this linear error contribution [3]

$$\sigma(\varepsilon_k^{\text{lin}}) = (k/N)^{1/2} \sigma(\delta B_\varepsilon)$$

$$= 2k^{1/2} N^{-1} (T_c / T_s + 1/\bar{n})^{1/2} \qquad (24)$$

where $\bar{n} = \langle n_m / T_s \rangle$. In figure 1 are plotted three experimental situations in a multiple tau correlator with three different relationships between the last delay time and a coherence time: $\tau_{\max} = T_c$, $\tau_{\max} = 10T_c$ and $\tau_{\max} = 100T_c$. The differences between standard, symmetric and compensated normalizations are evident.

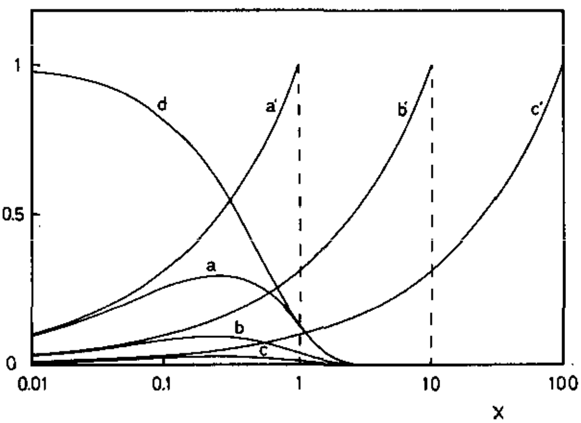

Figure 1. Three experimental situations in a multiple tau correlator with three different relationships between the last delay time and a coherence time: $x_{\max} = 1$ (a', a), $x_{\max} = 10$ (b', b) and $x_{\max} = 100$ (c', c). a', b', c': Relative linear error contributions $\varepsilon^{\text{lin}}(x)$ in the case of the standard normalization; a, b, c: the corresponding errors after using the symmetric normalization; d: autocorrelation function; $x = \tau / T_c$.

As can be concluded from figure 1, the compensated normalization is better than the symmetric one in these experimental cases (while using a 'multiple tau' correlator [1, 6, 8]) in which a measured delay time cannot be much longer than a coherence time, i.e. when a measurement time is limited or when a measured autocorrelation function has a long tail.

References

[1] Schätzel K, Drewel M and Stimac S 1988 *J. Mod. Opt.* **35** 711
[2] Saleh B E A and Cardoso M F 1973 *J. Phys. A: Math. Nucl. Gen.* **6** 1981
[3] Kojro Z 1990 *J. Phys. A: Math. Gen.* **23** 1363
[4] Degiorgio V and Lastovka J B 1971 *Phys. Rev.* A **4** 2033
[5] Jakeman E, Pike E R and Swain S 1971 *J. Phys. A: Gen. Phys.* **4** 517
[6] Schätzel K 1985 *Inst. Phys. Conf. Ser.* No. **77** 175–84
[7] Schätzel K 1983 *Opt. Acta* **30** 155–66
[8] Schulz-DuBois E O 1984 *Photon Correlation (lecture at Viana do Castelo, Portugal)*
[9] Wentzel M, Burchard W and Schätzel K 1986 *Polymer* **27** 195