Research Article

Exponentially Modified Peak Functions in Biomedical Sciences and Related Disciplines

A. Golubev

Saint-Petersburg State University, Saint Petersburg, Russia

Correspondence should be addressed to A. Golubev; lxglbv@rambler.ru

In many cases relevant to biomedicine, a variable time, which features a certain distribution, is required for objects of interest to pass from an initial to an intermediate state, out of which they exit at random to a final state. In such cases, the distribution of variable times between exiting the initial and entering the final state must conform to the convolution of the first distribution and a negative exponential distribution. A common example is the exponentially modified Gaussian (EMG), which is widely used in chromatography for peak analysis and is long known as ex-Gaussian in psychophysiology, where it is applied to times from stimulus to response. In molecular and cell biology, EMG, compared with commonly used simple distributions, such as lognormal, gamma, and Wald, provides better fits to the variabilities of times between consecutive cell divisions and transcriptional bursts and has more straightforwardly interpreted parameters. However, since the range of definition of the Gaussian component of EMG is unlimited, data approximation with EMG may extend to the negative domain. This extension may seem negligible when the coefficient of variance of the Gaussian component is small but becomes considerable when the coefficient increases. Therefore, although in many cases an EMG may be an acceptable approximation of data, an exponentially modified nonnegative peak function, such as gamma-distribution, can make more sense in physical terms. In the present short review, EMG and exponentially modified gamma-distribution (EMGD) are discussed with regard to their applicability to data on cell cycle, gene expression, physiological responses to stimuli, and other cases, some of which may be interpreted as decision-making. In practical fitting terms, EMG and EMGD are equivalent in outperforming other functions; however, when the coefficient of variance of the Gaussian component of EMG is greater than ca. 0.4, EMGD is preferable.

"Essentially, all models are wrong, but some are useful."
Box G. E. P. and Draper N. R. (1987), Empirical Model Building and Response Surfaces, John Wiley & Sons, NY, p. 424.

1. Introduction

The normal (Gaussian) and the exponential are probably the most widely known distributions and quite ubiquitous, too. No wonder that situations are possible where they are expected to meet each other. The resulting composite distributions have been suggested to be relevant, for example, to times between consecutive cell divisions [1–4] in cell biology and to times from stimulus to response [5–8] in psychophysiology.

Generally speaking, when the times of the passages of certain type objects from their initial to their final state are composed of variable times of their transition to an intermediate state and of their dwelling in the intermediate state, out of which they exit at random to their final state, then the random variable that represents the overall passage time of any such object is the sum of two independent random variables, the transit time and the dwell time, and the distribution of the overall passage times is defined as the convolution of the distributions of its summands [9]. The convolution of a Gaussian distribution and a negative exponential distribution is known as the exponentially modified Gaussian (EMG). Its generic plot is shown in Figure 1, and the approximations of empirical datasets with EMG are shown in Figures 2, 3, and 5.
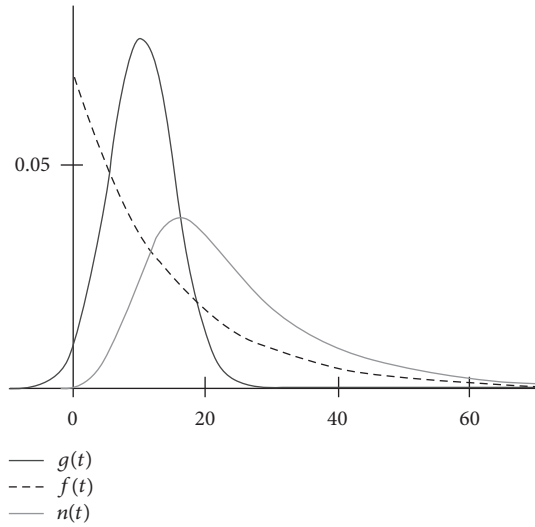
FIGURE 1: The plots of a Gaussian $g(t)$ and an exponential distribution $f(t)$ and of the result of their convolution $n(t)$ (EMG) at $S = 10$, $\sigma = 5$, and $k = 0.07$. The plot may also be regarded as showing the result of the deconvolution of an EMG into its exponential $f(t)$ and Gaussian $g(t)$ components. The Gaussian component may significantly extend to the negative domain even if $n(t)$ (EMG) may seem to be all nonnegative, which may be especially misleading if data points at small t, for example, within 5 time units, are lacking.
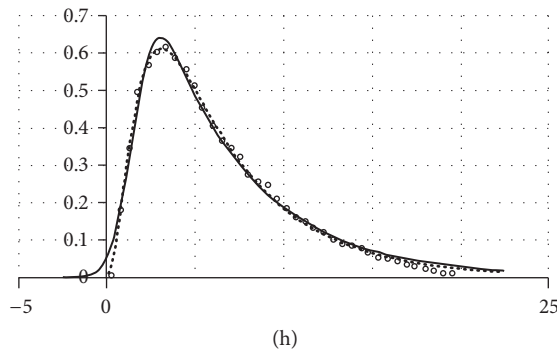

FIGURE 2: Data on intervals between the transcriptional bursts of the prolactin gene [19] approximated with EMG (solid line) and EMGD (dotted line). The data are shown with open circles. In the source paper [19], data were represented with histogram bars. The middle points of their tops are plotted here on an arbitrary vertical scale because only X-scale (time, h.) is what matters in the present context.

EMG has been introduced about 40 years ago in chromatography (see [10]) and psychophysiology (see [5]). In both cases, each accounting for about 100 entries in the Scopus database, EMG seems preferable over other skewed distributions, such as lognormal, gamma, Weibull, and Wald (inverse Gaussian), not only because of its better formal fits to data but, also and no less importantly, because of its straightforwardly interpretable parameters.
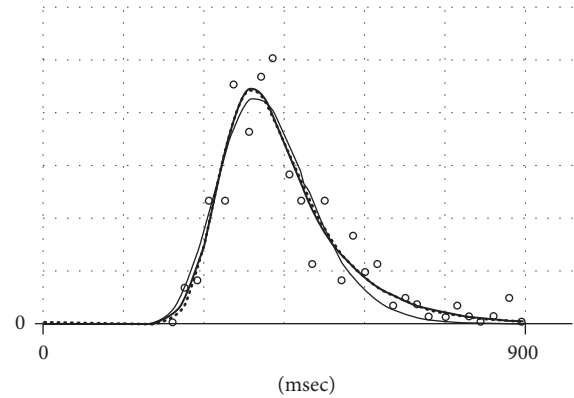

FIGURE 3: Data on response time distribution (RTD) picked up from [20] (open circles), which are approximated with EMG (thick solid line), EMGD (dots), and Wald distribution (thin solid line). The vertical scale is irrelevant.

More recently, EMG was suggested to be applicable to time distributions related to cell proliferation and differentiation [1, 2] and to distinct transcriptional states of active genes [2, 11]. Other novel EMG applications may be found in physiology [12], physics [13], and computer science [14].

However, there are cases when the deconvolution of an apparent EMG yields a Gaussian whose significant portion extends to the negative domain, which makes no physical sense. In such cases, nonnegative peak functions must be more appropriate for being convoluted with the negative exponent. In particular, a closed form expression for the exponentially modified gamma-distribution (EMGD) has been suggested and shown to be relevant to at least some of such cases [3]. In the present short review, the recent expansion of EMG applicability is considered in comparison with EMGD.

2. On Notations and Calculus

Exponentially modified functions result from applying the mathematical operator called convolution to nonmodified functions. The conventional definitions and notations for convolution are as follows:

$$
\begin{aligned}
y(x) = g(x) * f(x) &= \int_{-\infty}^{+\infty} f(x')\, g(x - x')\, dx' \\
&= \int_{-\infty}^{+\infty} f(x - x')\, g(x')\, dx',
\end{aligned}
\tag{1}
$$

where $g(x)$ is a nonmodified function; $f(x)$ is, in the present context, a negative exponential function (however, generally speaking, it may be any function); x' is an accessory (dummy) variable, which may be interpreted as the extent of shifting the plot of one function relative to the plot of the other function along the axis representing their common argument x; and $y(x)$ captures the resulting changes in the common area under the two plots.

When the domain of any of the convoluted functions is other than $(-\infty, +\infty)$, the limits of the convolution integral

must be changed, respectively. Being treated in different mathematical contexts, for example, [15–17], these relationships may be a source of confusion, especially with regard to integration limits, in using EMG in biomedicine and related fields, for example, chromatography (see below Section 3.1).

In probability theory, convolution is used to define the probability density function (PDF) of the sum h of two independent stochastic variables, say ξ_1 and ξ_2. In this case, according to Feller [9, p. 6],

$$h(s) = \int_{-\infty}^{+\infty} f(y) \, g(s - y) \, dy, \tag{2}$$

where f and g are the PDFs of ξ_1 and ξ_2, respectively, y is the value of the stochastic variable ξ_1, and s is the value of the sum of the stochastic variables $\xi_1 + \xi_2$.

Integration limits in (2) are $-\infty$ and $+\infty$. However, as pointed out, for example, in [9, p. 7] and [18, p. 216], integration limits must be different when at least one of the stochastic variables has nonzero probability densities only within a limited domain.

If probability densities of one of the variables are zero in the negative domain, then

$$h(s) = \int_{-\infty}^{s} f(y) \, g(s - y) \, dy. \tag{3}$$

If probability densities of both variables are zero in the negative domain, then [9]

$$h(s) = \int_{0}^{s} f(y) \, g(s - y) \, dy. \tag{4}$$

For the subsequent discussion, notations will be chosen to account for the biological meanings behind them. In particular, the stochastic variable is time t in most cases. Another point is that empirical distributions in biology are usually obtained starting from a counting procedure, for example, by counting of cells having intermitotic periods or of subjects showing response times within certain intervals of t. Therefore, the notation $n(t)$ is chosen for the respective PDFs:

$$n(t) = \lim_{\Delta t \to 0} \left[\frac{N(t) - N(t - \Delta t)}{\Delta t} \right], \tag{5}$$

where N is the normalized number of counts.

Using this notation, the convolution of a normal distribution and an exponential distribution is written as

$$n(t) = g(t) * f(t), \tag{6}$$

where t is time (in the present context); $n(t)$ is the probability density of events of interest; that is, $n(t)dt$ is the normalized number of objects that have experienced this event in the time interval $(t, t + dt)$, for example, the number of cells that, after having simultaneously emerged by the division of their mother cells, have divided within the time interval $(t, t + dt)$; $g(t) = (1/\sigma\sqrt{2\pi}) \times \exp(-(S - t)^2/2\sigma^2)$ is Gaussian probability density function, its domain being $(-\infty, +\infty)$, S being the mean of a Gaussian, and σ^2 being the dispersion of a

Gaussian; and $f(t) = k \exp(-kt)$ is an exponential probability density function, its domain being $(0, +\infty)$ and k being the rate parameter ($1/k$ is the mean of an exponential distribution and $1/k^2$ is its dispersion).

The computation of a closed form solution of $g(t) * f(t)$ using a dummy variable t' (to make one of the convoluted functions shifted relative to the other by t') starts with the integral:

$$n(t) = g(t) * f(t)$$
$$= \frac{k}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{(t - S - t')^2}{2\sigma^2} \right) \exp(-kt') \, dt'. \tag{7}$$

It may be solved by introducing the error function erf, which is not representable with elementary functions:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} \, dt. \tag{8}$$

The resulting closed form solution of (7) is

$$n(t) = \frac{k}{\sqrt{\pi}}$$
$$\cdot \exp\left[k(S - t) + \frac{(k\sigma)^2}{2} \right] \left[1 - \mathrm{erf}\left(\frac{S - t + k\sigma^2}{\sqrt{2}\sigma} \right) \right]. \tag{9}$$

A generic plot of EMG built according to (9) is shown in Figure 1, where the convoluted distributions are also shown. Notably, approximating an experimental dataset with (10) allows making estimates of the parameters of the convoluted functions, including their mean values, for example, mean transit time and mean dwell time (see Section 1).

The equivalents of (9) having multipliers to account for differences in peak height are inbuilt in curve-fitting software designed for chromatographic analysis, such as Fityk or SAS Fit, or for more general purposes, such as TableCuve2D, which is used throughout this paper.

A problem with (9) is that, strictly speaking, it is inapplicable to cases where negative time domains have no physical sense (in other words, above-zero $n(t)$ at $t = 0$ is physically impossible). This relates to almost any biologically relevant situation where measured t values are times between sequential events, for example, from stimulus to response (Section 3.2), from cell birth to cell division (Section 3.3), or from termination to resumption of transcription at a defined active gene (Section 3.4). Figure 2 illustrates this point using times from termination to resumption of transcription (see Section 3.4) as an example.

It has been suggested [3] that in such cases the gamma-distribution may be more appropriate for convolution with a negative exponential distribution. The result of the convolution may be termed as the exponentially modified gamma-distribution (EMGD).

The convolution integral for EMGD may be written as

$$n(t) = g(t) * f(t) = \int_0^t f(t - t') \cdot g(t') \cdot dt'$$

$$= \int_0^t \frac{e^{-(t-t')/\tau}}{\tau} \cdot \frac{t'^{c-1} e^{-t'/b}}{b^c \Gamma(c)} \cdot dt', \qquad (10)$$

where $g(t) = t^{c-1} e^{-t/b}/b^c \Gamma(c)$ is the gamma-distribution, whose domain is $(0, +\infty)$, $\Gamma(c) = \int_0^\infty t^{c-1} e^{-t} dt$ is the gamma-function, and $f(t)$ is the negative exponential function.

The gamma-distribution is a generalization of the Erlang distribution $g(t) = t^{c-1} e^{-t/b}/b^c (c-1)!$, which is essentially the convolution of c identical exponential distributions whose $1/k = b$. In the gamma-distribution, the values of c are not limited to integers.

The negative exponential component of (10) is expressed as $f(t) = (1/\tau) \exp(-(1/\tau)t)$, where the dimensionality of τ is the same as that of b. Both, τ and b, are interpreted here as the mean (characteristic) times of objects dwelling in states out of which they exit at random. EMGD can approximate this situation when $\tau \gg b_{max} \gg (b_{max} - b_{min})$. The rationale for this condition is discussed in [3].

The solution of the improper integral by (10) may be written as

$$n(t) = e^{-t/\tau} \cdot b^{-c} \cdot \left(\frac{\tau - b}{\tau \cdot b} \right)^{-c} \cdot [\tau \cdot \Gamma(c)]^{-1}$$

$$\cdot \left[\Gamma(c) - \Gamma\left(c, \frac{\tau - b}{\tau \cdot b} t \right) \right], \qquad (11)$$

where $\Gamma(c, ((\tau - b)/(\tau \cdot b))t) = \int_{((\tau-b)/(\tau \cdot b)) \cdot t}^\infty e^{-t} \cdot t^{c-1} \cdot dt$ is an upper (complementary) incomplete gamma-function.

Algorithms for numerical treatment of incomplete gamma-functions are included in standard curve-fitting software tools, such as TableCurve 2D used in the present paper and earlier [1–3].

Figure 2 illustrates the application of EMG and EMGD to fitting data on the duration of intervals between the transcriptional bursts of the prolactin gene.

Although fits to data according to the determination coefficient (r^2) are very good and virtually identical for EMGD (0.997) and EMG (0.993), EMG extends to the negative domain and if treated as a PDF, it will not be integrated to unity on the physically valid domain $(0, +\infty)$; therefore, EMGD is preferable for physical reasons even though there are cases when its fit to transcriptional data is poorer than that of EMG [3].

With all that, data points related to an empirical distribution may be, and often are, located on the time axis so far from its origin that EMG and EMGD become equivalent as analytical approximations. This is illustrated in Figure 3 where data on distribution of times from stimulus to response in humans are shown (see Section 3.2 for more details).

In Figure 3, EMG and EMGD look virtually identical. This is no surprise because the distance of EMGD peak from the origin is defined by the parameter c of the gamma-distribution. With increasing c, the shape of a gamma-distribution approaches that of a Gaussian (in conformance
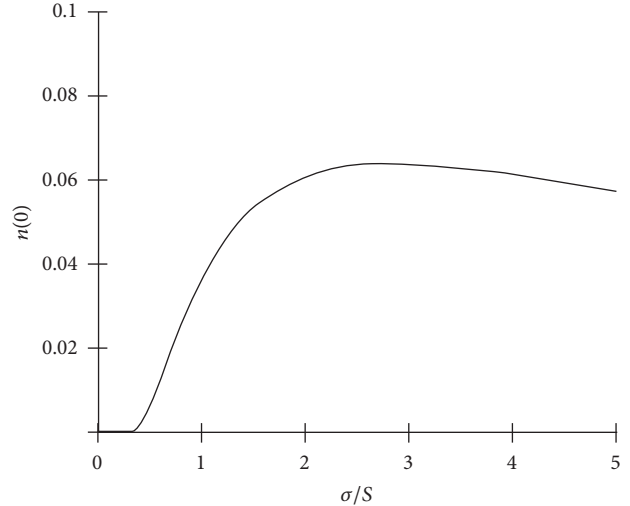

FIGURE 4: Increasing the height of the intersection of EMG (see (9)) with y-axis (n at $t = 0$) upon increasing the coefficient of variance (σ/S) of its Gaussian component. A greater height is associated with a more prominent extension of EMG to the negative domain.
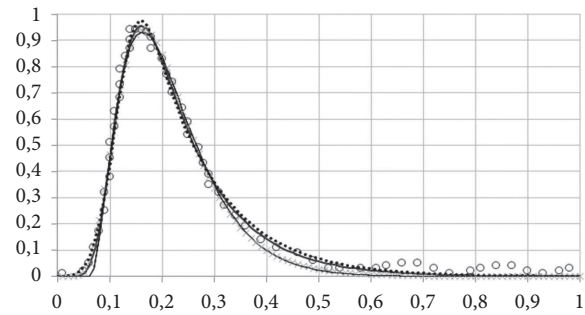

FIGURE 5: Fitting data on an amperometric spike produced by adrenaline release from adrenal medulla cell synaptic vesicle. Data points (open circles) are obtained by manual tracking of an amperometric trace presented in [36]. Scales are arbitrary. EMGD: thick solid line; EMG: dotted line; shifted gamma-distribution: thin solid line; lognormal distribution: symbols ×. Fits by determination coefficient (r^2): EMGD (0.993) > EMG (0.991) > lognormal = shifted gamma (0.990).

with the central limit theorem). However, with decreasing the distance, no part of EMGD will extend to the negative domain; and $n(t)$ at $t = 0$ will be zero. By contrast, with EMG, the interception of its plot with abscissa will increase with increasing the coefficient of variance of its Gaussian component, that is, the ratio of σ to S. The height of the interception of EMG plot with the ordinate will become appreciable at σ/S above 0.4 (Figure 4).

3. Applications and Implications of Exponentially Modified Peak Functions in Different Disciplines

3.1. Chromatography. In chromatography, EMG is employed since mid-1960s to describe chromatographic peaks, whose

shape is presumably determined by the diffusion-caused Gaussian blur of a compound during its passage through a column and by extracolumn effects of its exponential dilution in a detector cell [10, 21, 22]. The superposition of these two processes will skew an almost Gaussian symmetrical peak, whose width depends on diffusion, into a peak, whose skewness depends on detector cell volume. Importantly, conformance to EMG was demonstrated for the absorbance profiles of the bands of substances during their passage through a column. In this case, extracolumn effects are excluded, and EMG may be thought of as being generated by a combination of processes that include substance transfer (transit time) by the liquid phase and substance retention (dwell time) by the solid phase, from which the substance dissociates by a first-order process [23].

Different authors suggest a bewildering variety of mathematical expressions for EMG to be used in chromatography. The list compiled by Di Marco and Bombi [10] includes 18 very differently looking functions, among which 12 are judged as essentially correct and equivalent to each other, except for representing different steps towards the closed form solution of the convolution integral of EMG and/or implying different integration limits (see Section 2).

Hanggi and Carr [21] discussed potentially important problems arising upon different assumptions concerning the integration limits and suggested a solution for EMG where the lower limit is assumed to be zero and the upper limit to be the independent variable. The rationale of this assumption is that the domain of any component function of a convolution used in chromatography must be nonnegative, otherwise meaningless above-zero concentration values at elution time $t = 0$ may be yielded when integration limits $(-\infty, t)$ are used to compute EMG, and these values may become practically considerable at σ/S greater than about 0.4 (see Figure 4). However, both approaches to integration limits are equivalent, in practice, at low σ/S.

At present, (9) modified by a multiplier accounting for peak amplitude is used in commercial software packages, such as TableCurve2D, PeakFit, Fityk, and SASfit, devised for curve fitting or peak analysis.

3.2. Psychophysiology. The first attempts to use EMG in a biological context, which date back to early 1960s, relate to response time distributions (RTD) in psychophysiological studies (reviewed in [6, 20]). However, probably because the result of convolution of a Gaussian and a negative exponential function has been termed ex-Gaussian in this field, this approach has been developed independently from EMG usage in chromatography, which has started at about the same time. The idea that the positively skewed distributions of times from stimulus to response are consistent with the ex-Gaussian is based on the premise that times-to-response may include exponentially distributed periods needed for making a decision to respond and normally distributed periods needed to respond according to the decision (the alternative possibility, that is, that decision time is distributed normally and execution time is distributed exponentially, is also considered [5]). The belief that ex-Gaussian (=EMG) is appropriate to RTD is evidenced by the development of

dedicated software tools intended for finding best-fit EMG parameters of experimental RTD and for comparing EMG with other functions common in this field [20]. Remarkably, there has been no whatever crosstalk between chromatography and psychophysiology in designing software tools intended for very similar purposes.

The conformance of EMG and EMGD to an exemplary RTD presented in [20] is shown in Figure 3. The figure also shows an approximation of data with Wald (inverse Gaussian) distribution, which is also suggested for RTD data.

The use of EMG in psychophysiology was motivated by the intent to attribute a physiological significance to observed changes in RTD shapes. In particular, the centroid (S) was associated with the decision-making phase, whereas the rate constant (k) was associated with either the phase of execution of a decision or the phase or information processing required for a decision. The use of EMG was disputed by the proponents of another positively skewed PDF, that is, Wald (inverse Gaussian) distribution, based on other underlying premises rather than on a better or poorer fit (in the particular case shown in Figure 4, the fit of the Wald distribution is obviously inferior to those of EMG and EMGD). This controversy was discussed by Matzke and Wagenmakers [6] who concluded that EMG anyway remains a useful descriptive tool.

Some recent examples of using the deconvolution of EMG (ex-Gaussian) for distinguishing the different phases of psychophysical processes include cognitive changes that occur upon normal brain aging and Alzheimer disease [7] and in anxiety patients [24] and cocaine abusers [8] and variabilities found in word pronunciation time [25], eye fixation time during reading [26], and visual response time [5].

3.3. Cell Biology. It has been shown [1] that cell number distributions over times between sequential cell divisions (interdivision period, IDT) usually conform to EMG better than to traditionally used positively skewed distributions, such as lognormal or gamma. Subsequently, EMGD was shown to be no less appropriate as a descriptive model and more reasonable in biological (physical) terms, both EMG and EMGD outcompeting other positively skewed distributions [3].

The incentive to try EMG for fitting IDT was prompted by the transition probability model of cell cycle [27]. The model implies that, in a cycling cell population, daughter cells after having been born by their mother cell divisions will follow the courses of their cell cycles until reaching at different almost normally distributed times a state out which they exit at random; therefore, cell dwelling times in this state will be distributed exponentially. In a cell population, the exits of cells from this state will constitute a first-order process. If so, the overall IDT distribution in a cell population must be generated by the convolution of a Gaussian and a negative exponential function.

The intermediate state, which is mapped to G1 and is exited by cells at random, may be associated with the restriction point (RP) of cell cycle ([1, 2, 28] and references therein) and has been suggested to provide a time for cell fate to be determined, that is, whether a cell will pass its RP to

divide in its current state or will become committed, upon passing its RP, to change its state so as to make itself or its progeny different, up to the loss of the ability to divide, such as upon a terminal differentiation [1, 2].

EMGD has been shown to approximate IDT distribution as well as EMG does [3]. In technical terms, EMG may be more convenient for IDT decomposition into the probabilistic and the deterministic part of the cell cycle according to the transition probability model. However, EMGD is more appropriate in physically meaningful terms [3].

3.4. Molecular Biology. EMG has been suggested to be relevant to possible transcriptional mechanisms of the events that are associated with the restriction point of the cell cycle and can determine cell fate [2, 11].

The events are thought to result from discrete random changes in gene activity. The rationale for this suggestion is detailed elsewhere [2, 11, 29] based on published data about the direct observation of nascent mRNA association with cyclin D1 gene in single cells [30, 31] and on inferences from mathematical analysis of protein level fluctuations in single cells [19, 32, 33], in combination with some common-sense reasoning [29]. The available evidence reviewed in [3, 11, 33] supports the view that the activity of, at least, some genes consists of random bouts of their engagement in mRNA production, which alternate with periods of gene idleness occurring at gene-specific mean frequencies. All authors who address this issue agree that the engagement periods, whatever they are called, are distributed exponentially consistent with that they are terminated by the dissociation of transcription preinitiation complexes of transcription factors from gene promoters. The distribution of idleness periods was suggested [32, 33] to be generated by a process consisting of two first-order steps, therefore being consistent with the convolution of two exponential functions. Checking published distribution histograms for their fits to different functions shows that the EMG and EMGD provide better fits than the exponentially modified exponent (EME) or simple gamma-distribution does [3]. However, EMG almost invariably showed above-zero intercepts of plots with ordinates (see Figure 2), which is hard to interpret in physically meaningful terms.

The process of preparation of an idle gene to the next bout of its transcriptional engagement, which is switched on by the complete assembly of a proper preinitiation complex at gene promoter, is likely to involve more than two steps, and this will make several sequential first-order processes at a cell population level. Therefore, the overall time distribution must conform to the convolution of several exponents, which if the exponents are identical, is described with the Erlang distribution, a particular case of the gamma-distribution. When one of the processes stands out of all others in having a far lower rate constant, the overall distribution will be the convolution of an exponent, which is generated by this process, with the distribution collectively generated by all other processes (see Section 2 and [3]). When a gamma-distribution is used for an Erlang distribution in the convolution with an exponent, the parameter c of the gamma-function may prompt an estimate of the number of steps required to resume transcription after its interruption at an active gene [3].

The same must be true for the cell cycle where reaching an RP-associated state is likely to be brought about by a combination of numerous loosely correlated events, each constituting, at the cell population level, a process having a much higher rate compared with the process associated with cell passage through the RP.

In either case, EMG- or EMGD-based analysis of time distributions makes it possible to distinguish influences on the slowest process, such as cell passage through RP, from influences on the other constituents of the overall process. With regard to cell proliferation, the importance of such distinction follows from the possibility of the involvement of cell dwelling in the RP-associated state in cell differentiation [1]. This approach has also been proposed to distinguish between the modes of action of different anticancer drugs [4].

3.5. Other Cases of Applicability of Exponentially Modified Peak Functions. The conformance of a distribution to an exponentially modified peak function is an indication that the underlying processes collectively generate transit times, whose distribution is peaked, and exponentially distributed dwell times and that the mean dwell time is much longer than the mean characteristic time of any of the processes that determine the transit time. It has been shown [1] that EMG does not perform well when there are no reasons to expect that a distribution may be generated in that way. However, situations where such reasons do exist must be quite common. This ubiquity, which is not limited to the areas discussed above, is illustrated below with cases where fits to EMG and/or EMGD are observed outside the original fields of the applicability of these distributions.

In a study of microtubule growth in living cells by confocal fluorescence microscopy, the profiles of fluorescence intensity of complexes of microtubule-end binding protein with green fluorescent protein were fit to EMG based on the premise that these profiles are generated by a convolution of exponential decay with point Gaussian blur associated with microscopy [34]. Indeed, it may be confirmed that the profiles presented in [34] correspond to EMG better than to other positively skewed peak functions (not shown).

EMG was used to model human skin conductance changes in response to stimuli, such as noise or image [12]. Both EMG and EMGD may be found at the top of ranked lists of fits of different peak functions to skin conductance versus time plots (not shown).

In several studies of neurosecretion [35, 36], it was assumed that amperometric spikes, which result from the release of secreted substances from exocytotic vesicles, can be described by the convolution of a Gaussian with a decreasing exponential, which is partially governed by the diffusion of the substances towards the electrode. Figure 5 shows that such spikes may be fitted with EMG and EMGD equally well.

More remote to the initial field of EMG applicability are studies where EMG was used to model channel holding time distribution in public telephony systems [37] and packet payload lengths for two-player first-person shooter games in the server-to-client direction [14]. In physics, EMG emerged in an analysis of excited state lifetime distributions [13] and

was used to describe ion flight time distribution in a detector [38].

Searching for other cases where EMG may be applicable to time distributions for reasons discussed above shows that this may be true for some situation related to city traffic, such as that described in the paper [39] where the distribution of bus dwelling time at bus stops was fitted with a lognormal function; however, a better fit according to several goodness-of-fit criteria is obtained with EMG and EMGD (not shown).

4. Discussion

Making inferences from the shapes of distributions of variable parameters is a routine approach in physics; however, it has long been limited in biomedicine by enormous amounts of calculations. The state of things changes in recent years due to the availability of user-friendly software designed for this purpose.

A significant, in this regard, difference between biology and physics is that, in physics, distributions are often observed directly, whereas in biology they are generated by counting procedures, such as single-cell tracking, which may be very time-consuming and laborious. This problem will hopefully be ameliorated with the advent of novel automated single-cell tracking and other single-event monitoring techniques, which are increasingly introduced in biomedicine.

In the biological context, EMG and EMGD are useful as models, which suggest that, in a population of some objects there is, among the processes that generate the distribution of times required for each of the objects to pass between two observed consecutive states, a first-order process whose rate constant is much lower compared with those of all other processes. This slow process may be regarded as being generated by random discrete events that occur at a frequency making the mean interevent time comparable with the mean time between the two apparent consecutive states of the objects in question [1–3]. This interevent time is consistent with object dwelling in an unobservable intermediate state between the two observed states. Both, EMG and EMGD, make it possible to estimate the rate constant of this slow process, that is, the mean time of object dwelling in the intermediate state, and the contribution of all other processes to the overall time between the initial and final states and to its variance. Other similarly looking skewed distributions, such as lognormal or gamma, are not adequate to such cases.

However, no model can account for everything; therefore, any model is a compromise between physical (biological) relevance and mathematical tractability. To reiterate the epigraph to this writing: ". . . all models are wrong; the practical question is how wrong do they have to be to not be useful" (G. E. P. Box and N. R. Draper, 1987, p. 74). In this regard, EMG, which extends to the negative domain, is essentially wrong when it is applied to interevent time distributions; however, because of its easier mathematical tractability and the availability of ready-to-use curve-fitting tools, it is still useful as a descriptive option when the variation coefficient of its Gaussian component is low, within ca. 0.4. When the coefficient is higher than that, the use of EMGD is warranted. The failure of a distribution related to a class of phenomena, to which EMG or EMGD has been found to be generally applicable, to conform to them in a specific case may prompt that there is either a methodological bias or some unrecognized factor at work.

The use of EMG and/or EMGD in cell biology may help to supplement the assessment of the duration of different phases of cell cycle (G1, S, and G2) based on DNA content distributions with the assessment of the probabilistic and the deterministic part of cell cycle, according to the transition probability model, based on interdivision time distributions [1, 2]. This approach was used to discriminate between the modes of action of anticancer drugs [3, 4]. The importance of such discrimination follows from the suggestion that the rate of cell differentiation depends on the mean probabilistic rather than deterministic part duration, the former being associated with the restriction point of the cell cycle and suggested to provide cells with time for making a decision, by random choice, whether to continue proliferation or to become committed to withdrawal from the cell cycle [1].

Similarly, in psychophysiology, EMG or EMGD may help to distinguish between a phase required to make a decision to respond to a stimulus and a phase required to either execute the decision or to become prepared to make it [5, 6].

Based on the apparent pervasiveness of exponentially modified peak functions, such as EMG and EMGD, in different biological context—from molecular through cellular to physiological—it is tempting to suggest that this is a reflection of a common biological strategy of making decisions, for example, whether to commence or not to a next burst of gene transcription, cycle of cell proliferation, or response to a stimulus, each time there being alternative options chosen at random. As far as the probabilities of such choices depend on conditions, a population of biological systems can thus come to an optimal balance of different ways of coping with the unforeseeably changing world around them.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] A. Golubev, "Exponentially modified Gaussian (EMG) relevance to distributions related to cell proliferation and differentiation," *Journal of Theoretical Biology*, vol. 262, no. 2, pp. 257–266, 2010.

[2] A. Golubev, "Transition probability in cell proliferation, stochasticity in cell differentiation, and the restriction point of the cell cycle in one package," *Progress in Biophysics and Molecular Biology*, vol. 110, no. 1, pp. 87–96, 2012.

[3] A. Golubev, "Applications and implications of the exponentially modified gamma distribution as a model for time variabilities related to cell proliferation and gene expression," *Journal of Theoretical Biology*, vol. 393, pp. 203–217, 2016.

[4] D. R. Tyson, S. P. Garbett, P. L. Frick, and V. Quaranta, "Fractional proliferation: A method to deconvolve cell population dynamics from single-cell data," *Nature Methods*, vol. 9, no. 9, pp. 923–928, 2012.

[5] E. M. Palmer, T. S. Horowitz, A. Torralba, and J. M. Wolfe, "What Are the Shapes of Response Time Distributions in Visual

Search?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 1, pp. 58–71, 2011.

[6] D. Matzke and E.-J. Wagenmakers, "Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis," *Psychonomic Bulletin and Review*, vol. 16, no. 5, pp. 798–817, 2009.

[7] J. D. Jackson, D. A. Balota, J. M. Duchek, and D. Head, "White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease," *Neuropsychologia*, vol. 50, no. 3, pp. 357–366, 2012.

[8] S. Liu, S. D. Lane, J. M. Schmitz, C. E. Green, K. A. Cunningham, and F. G. Moeller, "Increased intra-individual reaction time variability in cocaine-dependent subjects: Role of cocaine-related cues," *Addictive Behaviors*, vol. 37, no. 2, pp. 193–197, 2012.

[9] W. Feller, "An Introduction to Probability Theory and Its Applications," *John Wiley & Sons*, 2008.

[10] V. B. Di Marco and G. G. Bombi, "Mathematical functions for the representation of chromatographic peaks," *Journal of Chromatography A*, vol. 931, no. 1-2, pp. 1–30, 2001.

[11] A. Golubev, "Genes at work in random bouts," *BioEssays*, vol. 34, no. 4, pp. 311–319, 2012.

[12] D. R. Bach, G. Flandin, K. J. Friston, and R. J. Dolan, "Modelling event-related skin conductance responses," *International Journal of Psychophysiology*, vol. 75, no. 3, pp. 349–356, 2010.

[13] B. Fournier and P. Coppens, "Measuring picosecond excited-state lifetimes at synchrotron sources," *Journal of Synchrotron Radiation*, vol. 19, no. 4, pp. 497–502, 2012.

[14] A. L. Cricenti and P. A. Branch, "The Ex-Gaussian distribution as a model of first-person shooter game traffic," *Multimedia Systems*, vol. 19, no. 3, pp. 221–229, 2013.

[15] R. E. Bellman and K. L. Cooke, *Differential-Difference Equations*, Academic Press, New York, NY, USA, 1963.

[16] V. Volterra, La Théorie Mathématique de la Lutte pour la Vie, et l'Expérience, 1936.

[17] F. G. Tricomi, Integral Equations, Courier Corporation, 1957.

[18] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury, Pacific Grove, Calif, USA, 2002.

[19] C. V. Harper, B. Finkenstädt, D. J. Woodcock et al., "Dynamic analysis of stochastic transcription cycles," *PLoS Biology*, vol. 9, no. 4, Article ID e1000607, 2011.

[20] A. Heathcote, "RTSYS: a DOS application for the analysis of reaction time data," *Behavior Research Methods, Instruments, & Computers*, vol. 28, no. 3, pp. 427–445, 1996.

[21] D. Hanggi and P. W. Carr, "Errors in exponentially modified Gaussian equations in the literature," *Analytical Chemistry*, vol. 57, no. 12, pp. 2394-2395, 1985.

[22] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, and P. Tikhonov, "Reconstruction of chromatographic peaks using the exponentially modified Gaussian function," *Journal of Chemometrics*, vol. 25, no. 7, pp. 352–356, 2011.

[23] A. M. Hupp and V. L. McGuffin, "Comparison of theoretical methods for extracting retention factors and rate constants in liquid chromatography," *Journal of Liquid Chromatography and Related Technologies*, vol. 33, no. 15, pp. 1427–1458, 2010.

[24] K. Bresin, M. D. Robinson, S. Ode, and C. Leth-Steensen, "Driven, distracted, or both? A performance-based ex-Gaussian analysis of individual differences in anxiety," *Journal of Personality*, vol. 79, no. 5, pp. 875–904, 2011.

[25] J. G. Holden and S. Rajaraman, "The self-organization of a spoken word," *Frontiers in Psychology*, vol. 3, Article ID 209, 2012.

[26] E. M. Reingold, E. D. Reichle, M. G. Glaholt, and H. Sheridan, "Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations," *Cognitive Psychology*, vol. 65, no. 2, pp. 177–206, 2012.

[27] J. A. Smith and L. Martin, "Do cells cycle?" *Proceedings of the National Academy of Sciences*, vol. 70, no. 4, pp. 1263–1267, 1973.

[28] A. Golubev, "Random discrete competing events vs. dynamic bistable switches in cell proliferation in differentiation," *Journal of Theoretical Biology*, vol. 267, no. 3, pp. 341–354, 2010.

[29] A. G. Golubev, "Accidental necessity, initiation of transcription, cell differentiation, and necessary accidents," *Biochemistry (Moscow)*, vol. 61, no. 7, pp. 928–938, 1996.

[30] S. Yunger, A. Kalo, P. Kafri et al., "Zooming in on single active genes in living mammalian cells," *Histochemistry and Cell Biology*, vol. 140, no. 1, pp. 71–79, 2013.

[31] S. Yunger, L. Rosenfeld, Y. Garini, and Y. Shav-Tal, "Single-allele analysis of transcription kinetics in living mammalian cells," *Nature Methods*, vol. 7, no. 8, pp. 631–633, 2010.

[32] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, "Mammalian genes are transcribed with widely different bursting kinetics," *Science*, vol. 332, no. 6028, pp. 472–474, 2011.

[33] D. M. Suter, N. Molina, F. Naef, and U. Schibler, "Origins and consequences of transcriptional discontinuity," *Current Opinion in Cell Biology*, vol. 23, no. 6, pp. 657–662, 2011.

[34] P. Kumar, M. S. Chimenti, H. Pemble et al., "Multisite phosphorylation disrupts arginine-glutamate salt bridge networks required for binding of cytoplasmic linker-associated protein 2 (CLASP2) to end-binding protein 1 (EB1)," *Journal of Biological Chemistry*, vol. 287, no. 21, pp. 17050–17064, 2012.

[35] J. D. Machado, A. Morales, J. F. Gomez, and R. Borges, "cAMP modulates exocytotic kinetics and increases quantal size in chromaffin cells," *Molecular Pharmacology*, vol. 60, no. 3, pp. 514–520, 2001.

[36] T. J. Schroeder, R. Borges, J. M. Finnegan, K. Pihel, C. Amatore, and R. M. Wightman, "Temporally resolved, independent stages of individual exocytotic secretion events," *Biophysical Journal*, vol. 70, no. 2 I, pp. 1061–1068, 1996.

[37] F. Barceló, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615–1625, 2000.

[38] P. Bochsler, L. Petersen, E. Möbius et al., "Estimation of the neon/oxygen abundance ratio at the heliospheric termination shock and in the local interstellar medium from IBEX observations," *Astrophysical Journal, Supplement Series*, vol. 198, no. 2, article no. 13, 2012.

[39] F. Li, Z. Duan, and D. Yang, "Dwell time estimation models for bus rapid transit stations," *Journal of Modern Transportation*, vol. 20, no. 3, pp. 168–177, 2012.