Statistical Science
2011, Vol. 26, No. 3, 352–368
DOI: 10.1214/11-STS364

A Problem in Particle Physics and Its Bayesian Analysis

Joshua Landon, Frank X. Lee and Nozer D. Singpurwalla

Abstract. There is a class of statistical problems that arises in several contexts, the Lattice QCD problem of particle physics being one that has attracted the most attention. In essence, the problem boils down to the estimation of an infinite number of parameters from a finite number of equations, each equation being an infinite sum of exponential functions. By introducing a latent parameter into the QCD system, we are able to identify a pattern which tantamounts to reducing the system to a telescopic series. A statistical model is then endowed on the series, and inference about the unknown parameters done via a Bayesian approach. A computationally intensive Markov Chain Monte Carlo (MCMC) algorithm is invoked to implement the approach. The algorithm shares some parallels with that used in the particle Kalman filter. The approach is validated against simulated as well as data generated by a physics code pertaining to the quark masses of protons. The value of our approach is that we are now able to answer questions that could not be readily answered using some standard approaches in particle physics.

The structure of the Lattice QCD equations is not unique to physics. Such architectures also appear in mathematical biology, nuclear magnetic imaging, network analysis, ultracentrifuge, and a host of other relaxation and time decay phenomena. Thus, the methodology of this paper should have an appeal that transcends the Lattice QCD scenario which motivated us.

The purpose of this paper is twofold. One is to draw attention to a class of problems in statistical estimation that has a broad appeal in science and engineering. The second is to outline some essentials of particle physics that give birth to the kind of problems considered here. It is because of the latter that the first few sections of this paper are devoted to an overview of particle physics, with the hope that more statisticians will be inspired to work in one of the most fundamental areas of scientific inquiry.

Key words and phrases: Exponential peeling, Markov chain Monte Carlo, mathematical biology, quarks, reliability, simulation, telescopic series.

1. INTRODUCTION AND OVERVIEW

Lattice Quantum Chromodynamics, or Lattice QCD, is an actively researched topic in particle physics. Many investigators in this field have received the Physics Nobel Prize, the 2004 prize going to Gross, Politzer and Wilczek, developers of the notion of "asymptotic freedom" that characterizes QCD. Underlying the Lattice QCD equations are issues of parameter estimation that have proved to be challenging. Essentially, one needs to estimate an infinite number of parameters from a finite number of equations, each

Joshua Landon is Assistant Professor, Department of Statistics, George Washington University, Washington, District of Columbia 20052, USA (e-mail: jlandon@gwu.edu). Frank X. Lee is Professor, Department of Physics, George Washington University, Washington, District of Columbia 20052, USA (e-mail: fxlee@gwu.edu). Nozer D. Singpurwalla is Professor, Department of Statistics, George Washington University, Washington, District of Columbia 20052, USA (e-mail: nozer@gwu.edu).

equation being an infinite sum of exponential functions.

The approach proposed here is Bayesian; it is driven by a computationally intensive Markov Chain Monte Carlo (MCMC) implementation. However, to invoke this approach, we need to introduce a latent parameter and then explore the "anatomy" of the QCD equations. This reveals a pattern, which when harnessed with some reasonable statistical assumptions provided a pathway to a solution. The inferences provided by our approach were successfully validated against simulated as well as real data. However, the real value of our approach is that it is able to answer questions that could not be answered using some of the conventional approaches of particle physics. The approach can therefore be seen as an addition to the lattice field theorists' data analysis tool kit.

The structure of the Lattice QCD equations is not as specialized as one is inclined to suppose. Indeed, such equations also appear in other contexts of engineering, physics, nuclear magnetic imaging and mathematical biology where they go under the label of "exponential peeling;" see Section 3.1. Our focus on the physics scenario is due to the fact that this is how we got exposed to the general problem addressed here.

This paper is directed toward both statisticians and physicists, and could serve as an example of the interplay between the two disciplines. The former may gain an added appreciation of problems in modern physics that can be addressed via statistical methods. In the sequel, they may also get to know more about particle physics and the beautiful theories about it that Mother Nature has revealed. It is, with the above in mind, that Section 2 is devoted to an overview of aspects of particle physics, its associated terminology and the awe inspiring discoveries about it. Reciprocally, the physicists may benefit by exposure to some modern statistical technologies that can be brought to bear for addressing problems that may have caused them some consternation.

Section 2 gives an overview of some essentials of particle physics, and the ensuing Lattice QCD equations. This section, written by a nonphysicist (NDS) but reviewed by a physicist (FXL), has been developed by fusing material from a variety of sources, some notable ones being Pagels (1982), Dzierba, Meyer and Swanson (2000), Yam (1993), Riordan and Zajc (2006) and Frank Wilczek's (2005) Nobel lecture. Interjected throughout this section are a few comments of historical interest; their purpose is to inform a nonphysicist reader about the individuals who have contributed to the building of a magnificent edifice. Section 2 concludes with a graphical display of the structure of matter via a template that is familiar to statisticians, in particular, those working in network theory and in reliability.

Section 3 pertains to an anatomy of the Lattice QCD equations and the resulting mathematical pattern that it spawns. It is not necessary to read Section 2 (save perhaps for an inspection of Figure 5) in order to read Section 3, which is where this paper really begins; indeed, Section 2 could have been delegated to an Appendix. Section 3 is a foundation for the rest of the paper. It is here that the inferential problem is introduced along with its accompanying notation and terminology. Section 3.1 gives a broad overview of the several other scenarios in science and engineering where the Lattice QCD type equations also arise. Of particular note are the several examples in mathematical biology wherein the QCD like equations are often discussed.

Section 4 pertains to the statistical model that the material of Section 3 creates, and an outline of the MCMC approach that is used to estimate the parameters of the model. These are the parameters that are of interest to physicists and other scientists. Section 5 pertains to validation against simulated and actual data and proof of principles. Section 6 pertains to some suggestions for extending the work done here, and strategies for overcoming some of the encountered difficulties. Section 7 concludes the paper.

Since the Lattice QCD equations can be seen as a prototype for similar equations that arise in other scientific endeavors, this paper also serves as an invitation to other statisticians to develop approaches for solving such equations using methods more sophisticated and/or alternate to the one we have entertained.

2. ESSENTIALS OF PARTICLE PHYSICS

The smallest quantity of anything we can see or feel is a *molecule*, and all matter is made up of molecules, which in turn are made up of *atoms*. Molecules and atoms are called *particles*, and the physics that describes the interactions between the particles is known as *particle physics*; see, for example, Griffiths (1987).

An atom consists of *electrons*, which carry a negative charge, and the electrons are centered around a *nucleus* that is made up of *protons* that carry a positive charge, and *neutrons* that carry no charge. Figure 1 illustrates the architecture of a carbon atom which has six electrons, six protons and six neutrons; it is denoted ${}^{12}_{6}\mathrm{C}$.

The protons and the neutrons are held together within the nucleus by a nuclear glue called the *pion*.

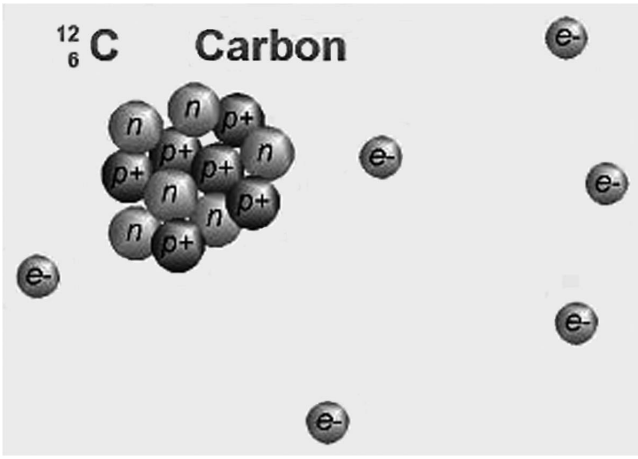


FIG. 1. *Architecture of a carbon atom.*

Similarly, the protons and the electrons are held together within the atom by a glue called the *photon*. The pions are said to be carriers (or mediators) of the *strong force* (or the *nuclear force*), and the photons are carriers of the *electromagnetic force*. Physicists look at the nuclear glues as force carrying particles, and thus collectively regard the electrons, the neutrons, the photons, the pions and the protons as *subatomic particles*. Figure 2 displays the structure of matter as understood around the 1946 time frame. The dotted lines of Figure 2 indicate the glued members.

In 1911, when Rutherford announced the structure of the atom, the existence of electrons and protons was known. The neutron, as a major constituent of the nucleus, was discovered in 1932 by Chadwick, and the pion was discovered in 1946. But these discoveries were just the tip of the iceberg. Many more subatomic particles have subsequently been discovered. Collectively, these subatomic particles are now called *hadrons*. Physicists speculate that there exist an infinite number of such hadrons. This discovery of hadrons was made possible by *accelerators*, which are essentially microscopes for matter.

The invention of the accelerators opened up the subnuclear world with the experimental discovery of thousands of new particles. The question thus arose as to what the hadrons could be saying about the ultimate structure of matter.

2.1 The Quark Structure of Matter

The current view is that hadrons are composite objects made out of more fundamental particles called *quarks*, and no one has ever seen a quark! This point of view came about in the early 1960s when Murray Gell-Mann discovered that the hadrons organized themselves into classes (or families) based on a mathematical symmetry. An easy way to understand why this organizational principle worked is to assume that the hadrons are made up of quarks, only three of which were needed to build the hadrons. These quarks were named the *up quark*, the *down quark* and the *strange quark*. For example, a proton has two up quarks and one down quark, whereas a neutron has two down quarks and one up quark. In general, every hadron is made up of quarks that orbit around each other in a specific configuration, each configuration resulting in a hadron. Figure 3 is an illustration of a quark orbit.

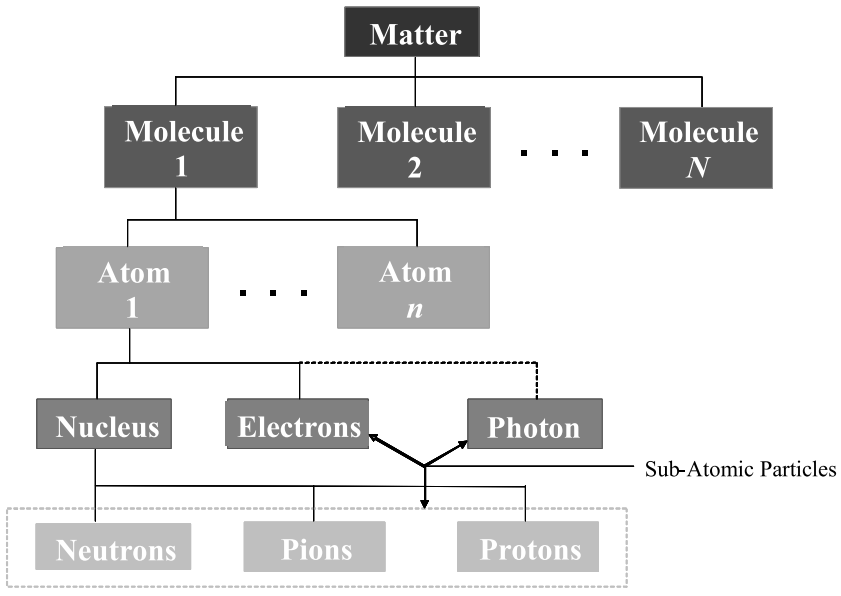


FIG. 2. *The structure of matter* (*circa* 1946).

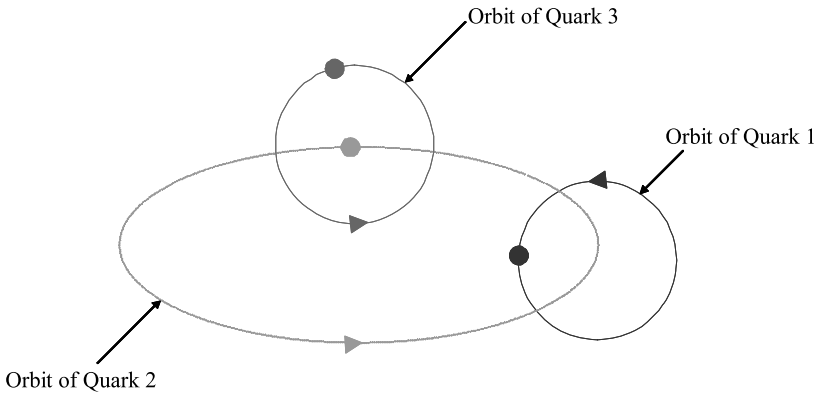


FIG. 3. *Illustration of a quark orbit.*

Since there could be several orbit configurations, there ought to be an infinite number of hadrons. The essence of Gell-Mann's idea is that hadrons are bound states of quarks, just like how the atoms are bound states of electrons, neutrons and protons. Furthermore, Gell-Mann postulated that there ought to exist a force carrying particle, called the *gluon*, that holds the quarks together. The gluon is said to be the carrier of the *strong force*. Figure 4 illustrates the quark structure of a hadron.

The quark model was purely a theoretical construct. Its validity was affirmed when Gell-Mann used it to postulate in 1962 the existence of a particle never seen before. This was a scientific breakthrough of the highest order! It showed that discoveries in physics can come from mathematical patterns—not just the laboratory. For unraveling the mathematical symmetries of the hadron, Gell-Mann received the 1969 Nobel Prize in Physics.

Figure 5 gives a pictorial representation of the quark structure of matter using a template that is familiar to statisticians. It represents an atom as a coherent (or logical) system with quarks as the basic building blocks of the system. The logic symbols of "and" and "or" are represented by and respectively. The neutrons and the protons can be regarded as subsystems, and the gluons, photons and the pions that link the quarks, the nucleus and the electrons can be seen as the structure (or link) functions of the system (cf. Barlow and Proschan, 1975). These are the carriers of the strong force and the electromagnetic force, respectively. Figure 5 contains Gell-Mann's famous quote that "everything that

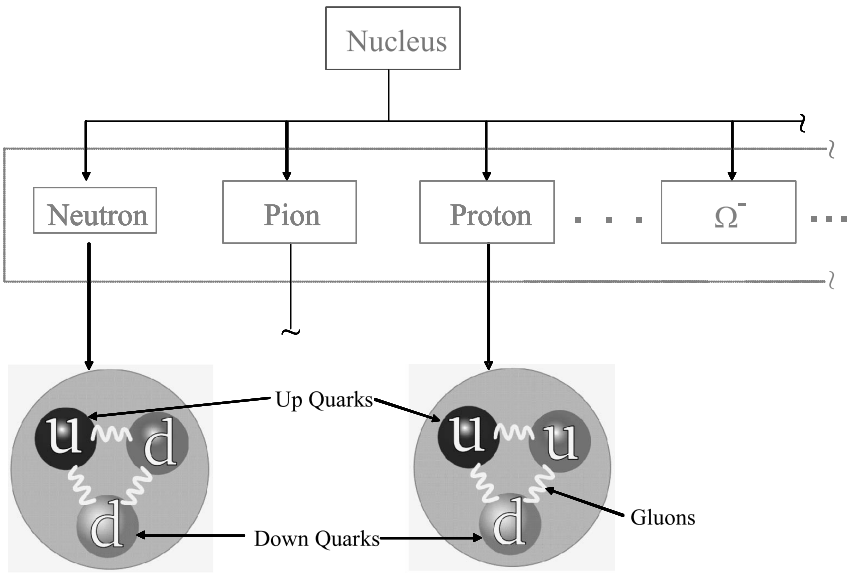


FIG. 4. *The quark structure of hadrons.*

"Everything that is not forbidden is compulsory"

Murray Gell-Mann

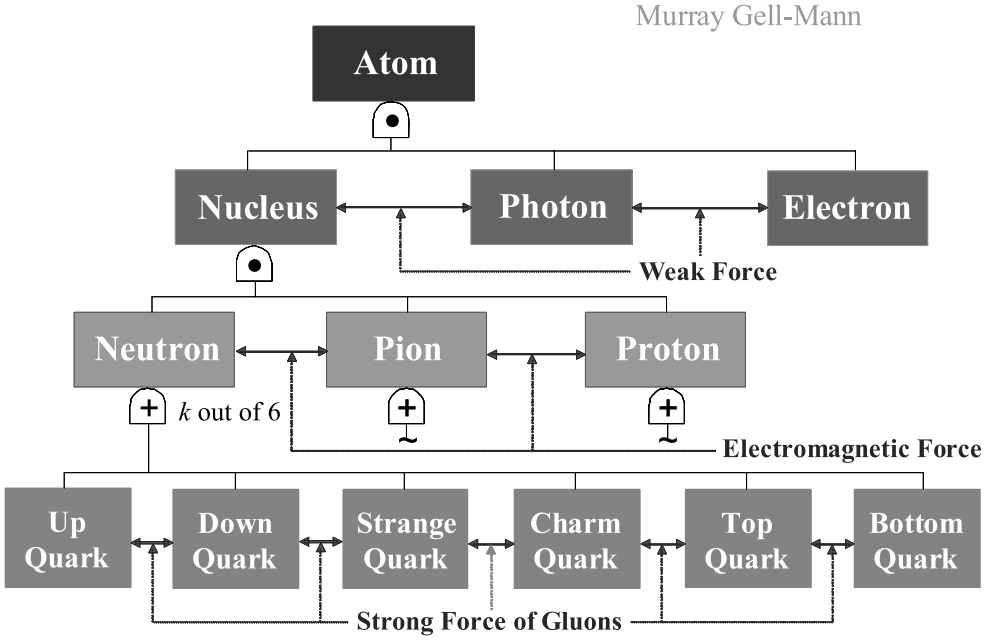


FIG. 5. *Matter as a coherent system.*

is not forbidden is compulsory;" the logical systems analogue to this quote is the notion of "irrelevance."

2.2 Quantum Chromodynamics and Lattice QCD

The theory of QCD can be thought of as a recipe for producing hadrons from quarks and gluons. Since quarks and gluons make up most of the known mass of the physical world, unraveling the quark structure of matter is the key to an understanding of the physical world, and thus the importance of the subject of this paper.

The QCD theory was successful in enunciating the properties of the hadrons. However, its complexity made its use for predicting unobservable quantum quantities, like quark masses, almost impossible. This is because solving the QCD equation (which is just one line) by analytical methods is difficult. The current approach is to solve the QCD equation numerically, by discretizing it over a space–time lattice. Lattice QCD refers to the representation of space–time as a scaffold in four dimensions wherein the quarks rest on the connecting sites, and the gluons as connections between the lattice points.

The scaffold is first restricted to a finite volume; it is then replicated with periodic boundary conditions. All this entails on the order of 100 million billion arithmetic operations on typical lattices; this is one example as to why physicists need supercomputers. Lattice QCD has been able to explain as to why a free quark has not been seen and will not be seen; this is because it will take an infinite amount of energy to isolate a quark.

Lattice QCD, being an approximation to the QCD, improves as the lattice points increase indefinitely and as the volume of the lattice grid expands. In so doing it opens up avenues for statistical methods to enter the picture. Physicists have explored some of these avenues, one of which is the focus of this paper; see Section 3 below.

3. THE UNDERLYING PROBLEM: QCD EQUATIONS

With Lattice QCD, an archetypal scenario is the estimation of an infinite number of parameters from a finite number of equations. The left-hand side of each equation is the result of a physics based Monte Carlo run, each run taking a long time to complete. Thus, there are only a finite number of runs. For example, a *meson correlator*, $G(t|\cdot)$, takes the form (cf. Lepage et al., 2002)

$$G(t|\cdot) = \sum_{n=1}^{\infty} A_n e^{-E_n t} \quad \text{for } t = 0, 1, 2, \ldots, \tag{3.1}$$

where the parameters A_n denote the *amplitude*, and E_n denote the *energy*. Also, $E_1 < E_2 < \cdots < E_n < \cdots$.

Interest centers around the estimation of A_n and E_n, $n = 1, 2, \ldots$, based on $G(t|\cdot)$, estimated as $\widehat{G}(t|\cdot)$, $t = 0, 1, \ldots, k$, for some finite k [23 in the case of Lepage et al. (2002)]. The physics codes which generate the $\widehat{G}(t|\cdot)$'s do not involve the A_n's and the E_n's, and are autocorrelated, thus the label "correlator." The physics codes also provide estimates of the autocorrelation matrix.

Deterministic approaches to solve for the A_n's and the E_n's cannot be invoked, and statistical approaches involving curve fitting by chi-square, maximum likelihood and empirical Bayes have proved to be unsatisfactory (cf. Morningstar, 2002). For an appreciation of these efforts, see Lepage et al. (2002), Fiebig (2002) and Chen et al. (2004); the latter authors propose what they call a "sequential empirical Bayes approach." However, empirical Bayes approaches use observed data to influence the choice of priors, and, as asserted by Morningstar (2002), are a violation of the Bayesian philosophy. Indeed, Fiebig (2002) states that "Bayesian inference has too long been ignored by the lattice community as an analysis tool. ... The method should be given serious consideration as an alternative for conventional ways."

Bayesian approaches alternate to ours have been considered by Nakahara, Asakawa and Hatsuda (1999). These authors entertain the use of maximum entropy priors, but, as claimed by Lepage et al. (2002), the accuracy of their estimator of E_2 is inferior to those obtained using other approaches. Because priors based on the principle of maximum entropy result in default priors, such priors also violate the Bayesian philosophy. The approach of Lepage et al. (2002) is Bayesian in the sense that prior information is used to augment a chi-square statistic which is then minimized. We find this work valuable because it articulates the underlying issues and provides a framework for examining the anatomy of the QCD equations, which enables us to identify a pattern, which in turn enables us to invoke the Bayesian approach we propose.

3.1 Relevance to Other Scenarios in Science and Engineering

The Lattice QCD architecture of equation (3.1) is not unique to physics. They occur in several other scenarios in the physical, the chemical, the engineering and the biological sciences, a few of which are highlighted below. Most attempts at estimation of the underlying parameters have involved least squares or numerical techniques based on local linearization with iterative improvements. Besides lacking a theoretical foundation vis-à-vis the requirement of coherence (cf. Bernardo and Smith, 1994, page 23), techniques have proved notoriously unreliable and not robust to slight changes in the experimental data (cf. Hildebrand, 1956).

Mathematical biology: *exponential peeling in compartment systems*. When considering radioactive tracers used for studying transfer rate of substances in living systems (cf. Robertson, 1957; Rubinow, 1975, page 125), sums of exponentials are encountered. Here, the $G(t|\cdot)$ of equation (3.1) represents the concentration of a substance, the t's are integer values of time, and the A_i's and the E_i's are constants that need to be estimated. Here interest generally centers around the case of $n = 2$, and the coefficients A_n and E_n of equation (3.1) are negative. An ad hoc graphical procedure called *the method of exponential peeling* is used to estimate the parameters (cf. Smith and Morales, 1944, Perl, 1960; van Liew, 1967).

Some other scenarios in biology where the Lattice QCD type equations appear are in bone metabolism studies and cerebral blood flow (cf. Glass and de Garreta, 1967), and in biological decay (cf. Foss, 1969). In the latter context, Dyson and Isenberg (1971) consider for fluorescence decay an equation of the type

$$y(t) = \sum_{j=1}^{m} \alpha_j \exp(-t/\tau_j), \quad 0 \le t \le T,$$

where $y(t)$ represents "moments of the fluorescence," α_j's the amplitudes [the A_n's of equation (3.1)], and the τ_j's are time constants corresponding to the E_n's of equation (3.1). Here the α_j's are zero for $j \ge m + 1$.

Gene expression data. When considering a time series of gene expression data (cf. Giurcaneanu et al., 2005), a system of equations paralleling that of equation (3.1) arises again. In this context $G(t|\cdot)$ represents "mRNA concentrations" as a function of time, and the parameters A_n and E_n describe interactions between the genes. In the gene expression context, as in the Lattice QCD context, the parameters E_n are increasing in n.

Nuclear magnetic resonance (*NMR*). NMR experiments often generate data that are modeled as the sum of exponentials (cf. Bretthorst et al., 2005). Experiments relying on NMR to probe reaction kineticis, diffusion, molecular dynamics and xenobiotic metabolism are some of the applications where parameter estimates provide insight into chemical and

biological processes. See, for example, Paluszny et al. (2008/09) who study brain tissue segmentation from NMR data.

Here one considers equations of the type

$$d_i = C + \sum_{j=1}^{m} A_j \exp\{-\alpha_j t_i\} + n_i,$$

where m is the number of exponentials and d_i a data value sampled at t_i. The parameters of interest are the decay rate constants α_j, the amplitudes A_j and the constant offset C. The n_s's are the error terms.

Electromechanical oscillations in power systems. Equations entailing the sum of exponentials are also encountered in the context of low frequency electromechanical oscillations of interconnected power systems, the impulse response of linear systems in networks, ultracentrifuge and a host of other relaxation and time-decay phenomena (cf. Dyson and Isenberg, 1971). For example, in the electromagnetic oscillations scenario, Sanchez-Gasca and Chow (1999) encounter an equation analogous to our equation (3.1) with $G(t|\cdot)$ denoting a signal and A_n connoting a signal residue associated with the "mode" E_n.

To summarize, the relationships of the type given by equation (3.1) arise in so many contexts of science and engineering that it seems to be quintessential, and almost some kind of law of nature. The Lattice QCD problem considered here can therefore be seen as a prototype and a convenient platform to exposit a statistical problem of general applicability. In most of the application scenarios described above, statistical methods have been used, many ad hoc, some empirical Bayesian and a few Bayesian (under the rubric of maximum entropy). Many of these methods have not exploited an underlying telescopic pattern in these equations which makes an appearance when a latent parameter is introduced into the system, and inference about the latent parameter made.

3.2 Anatomy of the Lattice QCD equations

An examination of equation (3.1) yields the following boundary conditions. $G(0|\cdot) = \sum_{n=1}^{\infty} A_n$, implying that the A_n's are constrained. When $t \to \infty$, $G(t|\cdot) = 0$, which implies that for large values of t, A_n and the E_n cannot be individually estimated. Thus, simulating $G(t|\cdot)$ for large t does not have a payback; consequently, it is futile to do such a simulation.

Since the E_n's increase with n, we may, as a start, reparameterize the E_n's as $E_n - E_{n-1} = c$, for some *unknown* c, $c > 0$, for $n = 2, 3, \ldots$. It will be argued later, in Section 6.1, that c is a latent parameter. Thus,

$$E_n = E_1 + (n-1)c, \quad n = 2, 3, \ldots \tag{3.2}$$

with E_1 and c unknown. With the above assumption in place, a parsimonious version of the Lattice QCD equation takes the form

$$G(t|\cdot) = e^{-E_1 t} \sum_{n=1}^{\infty} A_n e^{-(n-1)ct}, \quad t = 0, 1, 2, \ldots. \tag{3.3}$$

With c fixed, the parsimonious model given above reveals the following features:

(a) When t is *small*, the number of A_n's entering equation (3.3) is large; indeed, infinite when $t = 0$.

(b) When t is *large*, the number of A_n's we need to consider is small, because the combination of a large t with any n will make the term $A_n \exp(-(n-1)ct)$ get small enough to be ignored.

(c) Moderate values of t and n will also make the above term small, causing A_n to be irrelevant.

Figure 6 illustrates the feature that as t gets large, the number of A_n's one needs to consider gets small.

As a consequence of the above, for any fixed c, we can find a t_1 such that in the expression

$$e^{-tE_1}[A_1 + A_2 e^{-ct} + A_3 e^{-2ct} + \cdots + A_n e^{-(n-1)ct} + \cdots],$$

all the terms, save for A_1, are essentially zero.

Similarly, we can find a t_2, $t_2 < t_1$, such that all the terms save for A_1 and $A_2 e^{-ct_2}$ get annihilated. Continuing in this vein, there exists a sequence $t_k < t_{k-1} < \cdots < t_2 < t_1$, such that all that matters are the terms

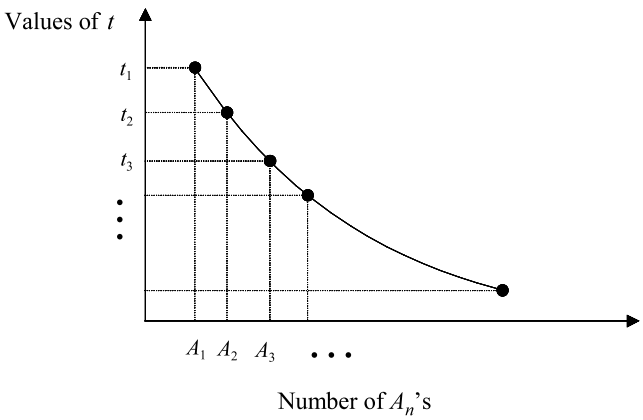


FIG. 6. *Number of A_n's as a function of t.*

associated with $A_1, A_2, \ldots, A_k$. In what follows, we suppose that k is *specified*.

Thus, for any fixed c and k, with $t_1 > t_2 > \cdots > t_k$ chosen in the manner described above, our parsimonious version of the Lattice QCD equations *telescope* as follows:

$$\begin{aligned} G(t_1|\cdot) &= e^{-E_1 t_1} A_1, \\ G(t_2|\cdot) &= e^{-E_1 t_2}(A_1 + A_2 e^{-ct_2}), \\ G(t_3|\cdot) &= e^{-E_1 t_3}(A_1 + A_2 e^{-ct_3} + A_3 e^{-2ct_3}), \\ &\vdots \\ G(t_k|\cdot) &= e^{-E_1 t_n}\big(A_1 + A_2 e^{-ct_k} + \cdots \\ &\qquad + A_k e^{-(k-1)ct_k}\big). \end{aligned} \tag{3.4}$$

To summarize, by introducing the constant c, fixing a k, and identifying an underlying pattern in the Lattice QCD equations, we have reduced the problem to the case of k equations and $(k+2)$ unknowns, $A_1, \ldots, A_k, E_1$ and c. The choice of what k to choose is determined by the number of physics code based estimates $\widehat{G}(t), t = 0, 1, \ldots, k$, that can be done and are available.

4. STATISTICAL MODEL: SOLVING THE QCD EQUATIONS

Many have expressed the view that it would be considered good progress if trustworthy estimates of just A_1, A_2, E_1 and E_2 can be had. The other pairs (A_3, E_3), $(A_4, E_4), \ldots,$ can be considered later; see Section 6. Thus, we start by focusing attention on the first two equalities of equation (3.4); that is, the case $k = 2$ and some fixed c. Specifically, we consider

$$\begin{aligned} G(t_1|A_1, E_1) &= e^{-E_1 t_1} A_1 \quad \text{and} \\ G(t_2|A_1, E_1, A_2, c) &= e^{-E_1 t_2}(A_1 + A_2 e^{-ct_2}). \end{aligned} \tag{4.1}$$

If $y_i = \widehat{G}(t_i|\cdot), i = 1, 2$, denotes the physics code based evaluations of $G(t_i|\cdot)$, then our aim is to estimate A_1, E_1, A_2 and c, in light of y_1 and y_2. To set up our likelihoods, we take a lead from what has been done by Nakahara, Asakawa and Hatsuda (1999), and by Lepage et al. (2002), to write

$$\begin{aligned} Y_1 &= G(t_1|\cdot) + \varepsilon_1 \quad \text{and} \\ Y_2 &= G(t_2|\cdot) + \varepsilon_2, \end{aligned} \tag{4.2}$$

where $\varepsilon_i \sim N(0, \sigma_i^2), i = 1, 2$, and $\mathrm{Corr}(\varepsilon_1, \varepsilon_2) = \rho_{12}$.

Besides providing y_1 and y_2, the physics codes also provide σ_1^2, σ_2^2 and ρ_{12}. As a consequence, the statistical model boils down to the bivariate normal distribution,

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} G(t_1) \\ G(t_2) \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right). \tag{4.3}$$

Writing out a likelihood function for the unknowns A_1, E_1, A_2 and c, based on equation (4.3), is a straightforward matter. However, we need to bear in mind that since the parameters A_1 and E_1 appear in both $G(t_1|\cdot)$ and $G(t_2|\cdot)$, both y_1 and y_2 provide information about A_1 and E_1, with y_2 providing information about A_2 and c as well. To exploit this feature, we construct our likelihoods based on the marginal distribution of Y_1, and the conditional distribution of Y_2 given Y_1. That is, on

$$Y_1 \sim N(A_1 e^{-E_1 t_1}, \sigma_1^2) \tag{4.4}$$

and

$$\begin{aligned} (Y_2|Y_1 = y_1) \sim N\Big(& G(t_2|\cdot) + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - G(t_1|\cdot)), \\ & \sigma_2^2(1 - \rho_{12}^2)\Big). \end{aligned} \tag{4.5}$$

Specifically, the likelihood of A_1 and E_1, with y_1 fixed, is

$$\mathcal{L}(A_1, E_1; y_1) \propto \exp\left[-\frac{(y_1 - A_1 e^{-E_1 t_1})^2}{2\sigma_1^2}\right], \tag{4.6}$$

and the likelihood of A_1, E_1, A_2 and c, with y_2 fixed, and the effect of y_1 incorporated via the posterior distribution of A_1 and E_1, is of the form

$$\begin{aligned} &\mathcal{L}(A_1, E_1, A_2, c; y_1, y_2) \\ &\quad \propto \exp\Bigg[-\Big\{y_2 - \big(e^{-E_1 t_2}(A_1 + A_2 e^{-ct_2})\big) \\ &\qquad + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_1 - A_1 e^{-E_1 t_1})\Big\}^2 \\ &\qquad \cdot [2\sigma_2^2(1 - \rho_{12}^2)]^{-1}\Bigg]. \end{aligned} \tag{4.7}$$

In the above development, the covariance matrix is provided by the physics code. As suggested by a referee, a deeper investigation of this matrix may be called for, because with increasing t, the variances are likely to increase, posing computational challenges to the proposed approach.

4.1 Specification of the Prior Distributions

To implement our Bayesian approach, we need to make assumptions about conditional independence, and assign prior distributions for the unknown parameters. The priors that we end up choosing in Section 5 are not based on knowledge of the underlying physics, but are proper priors based on an appreciation of the material in Morningstar (2002), Lepage et al. (2002) and Fleming (2005).

The A_i's are supposedly between 0 and 1, and no relationship between them has been claimed. Thus, it is natural to assume that A_1 and A_2 are apriori independent, and have a beta distribution on (0, 1) with parameters (α, β); we denote this as $\mathcal{B}(A_i; \alpha, \beta), i = 1, 2$. The relationship between E_1 and c is less straightforward. We conjecture that the larger the E_1, the smaller the c, and that E_1 can take values over $(0, \infty)$. It is therefore reasonable to assume that the prior on E_1 is a gamma distribution with scale parameter η and shape parameter λ; we denote this by $\mathcal{G}(E_1; \eta, \lambda)$. Some other meaningful choices for a prior on E_1 could be a Weibull, or a Pareto, the latter being noteworthy as a fat-tailed distribution. To encapsulate the dependence between E_1 and c, we suppose that, given E_1, c has a uniform distribution over $(0, \omega/E_1)$, for some $\omega > 0$. Finally, we also assume that E_1 and c are independent of all the A_i's.

The above choice of priors, with user specified hyperparameters α, β, ω, λ and η, is illustrative. In principle, any collection of meaningful priors can be used, since the ensuing inference is done numerically via a Markov chain Monte Carlo (MCMC) approach.

Lepage et al. (2002), and also Morningstar (2002), seem to use independent Gaussian priors for the parameters in question—see equations (8) and (11) respectively. Indeed, Morningstar (2002) makes the claim that "practitioners often restrict the choice of a prior to some familiar distributional form." The restricted parameter space makes the choice of Gaussian priors questionable. An overview of how the MCMC is invoked here is given next.

4.2 An Outline of the MCMC Excercise

The telescopic nature of the Lattice QCD equations suggests that the MCMC will have to be conducted in the following three phases:

Phase I. Using $E_1^{(0)}$ as a starting value and y_1 as data, obtain the posterior distribution of A_1 and E_1 via equation (4.6) as the likelihood, and 1,000 iterations of the MCMC run.

Phase II. Using $c^{(0)}$ as a starting value, and y_2 as data, obtain a sample from the posterior distribution of A_1, E_1, A_2 and c via the likelihood of equation (4.7), and 1,000 iterations of the MCMC run. Sample values of A_1 and E_1 from their posterior distributions obtained in Phase I will serve as the priors of A_1 and E_1 in Phase II. Since the parameters A_1 and E_1 reappear in the likelihood of equation (4.7) as the mean of y_2, Phase II of the MCMC run captures the effect of y_2 on these parameters. The effect of y_1 was captured in Phase I.

Phase III. Repeat Phase I and Phase II m times using new starting values of E_1 and c to produce a sample of size m from the posterior distribution of A_1, E_1, A_2 and c, with y_1 and y_2 as the data.

The MCMC exercise described above is routine, but computer intensive and entails 12 steps, six in each phase, and this too for a highly curtailed version of the Lattice QCD equations. The details of how this is done could be interesting, because they involve some discretization of the simulated posterior distributions, and working with individual sampled values reminiscent of that done in particle Kalman filtering (cf. Gordon, Salmond and Smith, 1993). Thus, we label our approach as *Particle MCMC*. More details are given in Landon (2007), and the method illustrated in the Appendix. The software can be downloaded at http://www.gwu.edu/~stat/irra/Lattice_QCD.htm.

4.3 A Caveat of the Proposed Scheme

The caveat mentioned here stems from the features that c has been fixed, and that the MCMC runs are centered around fixed values of y_1 and y_2. To see why, recall that our parsimonious version of the Lattice QCD equations [see equation (3.4)] is based on those t_i's for which the exponential terms vanish; however, the t_i's are determined by a fixed value of c. Thus, any change in the value of c will bring about a change in the values of t_i, and, as a consequence, the Lattice QCD equations will also have to be different. This would be tantamount to obtaining new values of the y_i's. However, all the likelihoods in the MCMC runs are based on fixed values of the y_i's; see equations (4.6) and (4.7). But a change in the value of c is inevitable, because in Phase II of the MCMC run one iterates around sampled values from the posterior distribution of c, so that the initial $c^{(0)}$ systematically gets replaced by $c^{(1)}$, $c^{(2)}, \ldots, c^{(1,000)}$.

A way to overcome this caveat is to recognize that for any $c^{(i)} > c^{(0)}$, $i = 1, 2, \ldots$, the exponential terms mentioned above will continue to vanish, so that any

specified values of y_i will continue to satisfy the right-hand side of equation (4.2).

A strategy to ensure that the successively generated values of $c^{(i)}$, $i = 1, 2, \ldots$, will tend to be greater than $c^{(0)}$ is to pick small values of $c^{(0)}$ for each of the m iterations of Phase III of the MCMC algorithm. During the course of the MCMC runs, should one encounter a generated value of $c^{(i)}$ that is smaller than $c^{(0)}$, then one should *discard* the so-generated value $c^{(i)}$, and generate another value of $c^{(i)}$. Hopefully, the number of discarded $c^{(i)}$'s will not be excessive, but if they are, then the starting value $c^{(0)}$ should be decreased, and new values of t_1 and t_2 obtained. This of course would be tantamount to obtaining new values of y_1 and y_2 as well.

5. PROOF OF PRINCIPLE: VALIDATION AGAINST DATA

We first validate the accuracy of our approach against simulated data. For this, we choose $A_1 = 0.8$, $A_2 = 0.6$, $A_3 = 0.4$, $A_4 = 0.2$, $A_5 = 0.1$, and $A_i = 0$ for $i \geq 6$. We also choose $E_1 = 0.9$ and $c = 0.5$. Using these values in equation (3.3), we compute $G(t)$, for $t = 1, 2, \ldots, 12$; these are shown in column 3 of Table 1. Since $Y_t = G(t|\cdot) + \varepsilon_t$, with $\varepsilon_t \sim N(0, \sigma_t^2)$ [see equation (4.2)], we generate $y_1, \ldots, y_{12}$, assuming the ε_t's are independent, with $\sigma_t = 0.001 \times G(t) \times t$; these are shown in column 4 of Table 1. We next identify those t's for which the leading exponential terms vanish. These happen to be t_1 at $t = 12$, t_2 at $t = 6$, and t_3 at $t = 4$; see column 2 of Table 1. Our aim is to invoke the methods of Section 4 on the entries of Table 1, to see if the constants specified above can be returned.

TABLE 1
Simulated data for validating approach

Time t	Index t_i	$G(t)$	y_i
1		0.54874373	0.54900146
2		0.17764687	0.17756522
3		0.06387622	0.06373037
4	t_3	0.02422158	0.02414992
5		0.00945326	0.00952723
6	t_2	0.00375071	0.00377058
7		0.00151265	0.00151498
8		0.00060552	0.00061147
9		0.00024486	0.00024698
10		0.00009923	0.00009821
11		0.00004026	0.00004007
12	t_1	0.00001635	0.00001625

With the above in place, Phases I, II and III of the MCMC run were made arbitrarily choosing the hyper-parameters $\alpha = \beta = \eta = \lambda = \omega = 1$, and $m = 1{,}000$.

5.1 Results Based on Simulated Data

Figure 7(a) and (b) shows the posterior distributions of E_1 based on y_{12}, and on y_{12} and y_6, respectively. Recall that y_{12} corresponds to t_1, and y_6 corresponds to t_2. Note that the posterior distribution of Figure 7(a) becomes the prior distribution for the construction of the posterior distribution of Figure 7(b). Both the distributions of Figure 7 indicate a modal value of 0.9, suggesting a tendency to converge to the true value of E_1. Furthermore, the difference between the two distributions is not very great, suggesting that y_6 may not be contributing much toward inference for E_1, beyond that provided by y_{12}.

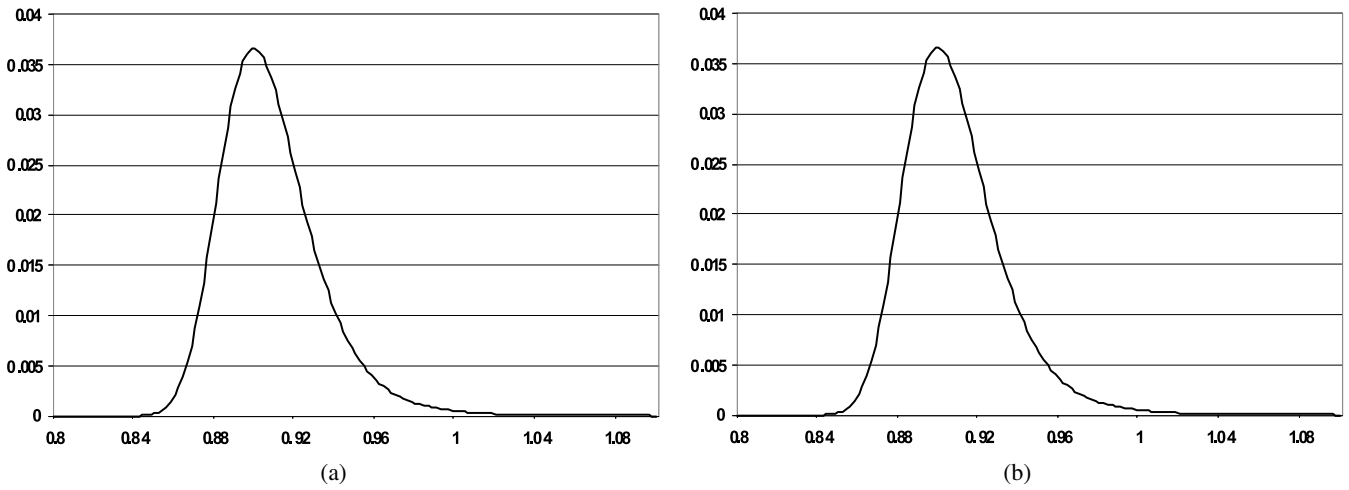


FIG. 7. *Posterior distribution of* E_1. (a) *Posterior of* E_1 *based on* y_{12}. (b) *Posterior of* E_1 *based on* y_{12} *and* y_6.

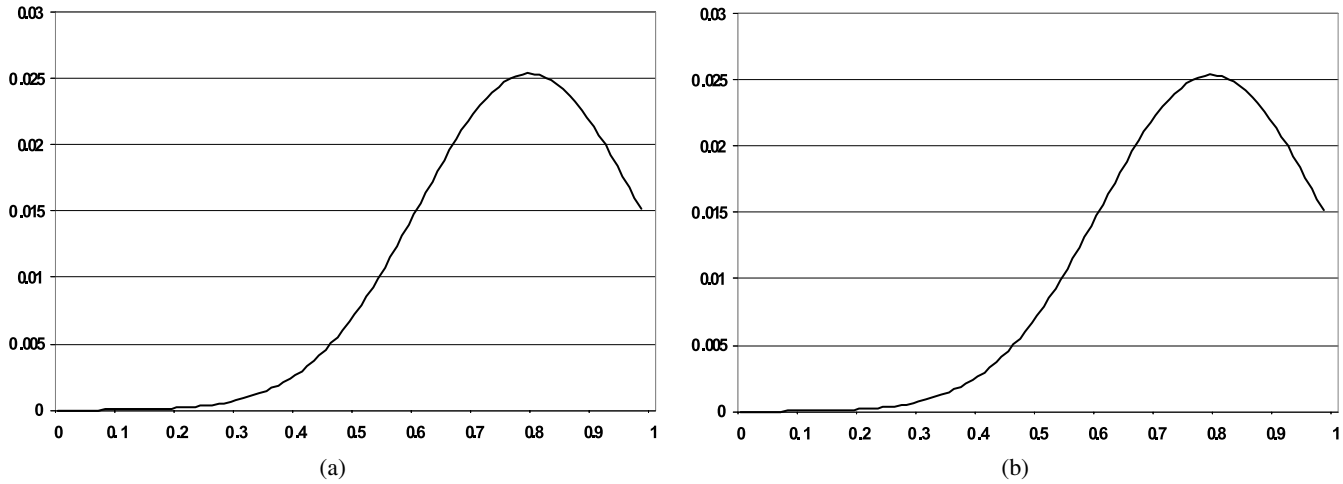


FIG. 8. *Posterior distribution of* A_1. (a) *Posterior of* A_1 *based on* y_{12}. (b) *Posterior of* A_1 *based on* y_{12} *and* y_6.

A similar feature is revealed by the posterior distributions of A_1, shown in Figures 8(a) and (b). These distributions have a modal value of 0.8, suggesting again a convergence to the true value of A_1.

Figures 9 and 10 show the posterior distributions of A_2 and c, based on y_{12} and y_6. Their modal values of 0.6 and 0.5 suggest convergence of the posteriors to their true values. Thus, based on this simulation exercise, we may claim that, despite an arbitrary choice of hyperparameters, the proposed MCMC procedure is able to show recovery of the input values of A_1, E_1, A_2 and c to a meaningful degree of accuracy.

5.1.1 *Sensitivity of posteriors to priors.* In this section we explore the sensitivity of the posterior distributions of A_1, E_1, A_2 and c when the hyperparameters of their prior distributions vary. We also explore the effect of using a thick-tailed prior distribution for E_1, in particular, a Pareto distribution, instead of the gamma distribution used before.

Figure 11 shows the posterior distributions of E_1 for different values of the scale λ and shape η parameters of its gamma prior. Verify that the posterior distributions get centered around its true value of 0.9 even when the prior mean is as large as 10. The values of the chosen hyperparameters are indicated in the legend accompanying Figure 11.

In Figures 12 and 13 we show the posterior distributions of A_1 and A_2 for different values of the hyperparameters α and β; see the legend accompanying these figures. Whereas the posterior distribution of A_2 appears to be very robust against the various choices for its prior distributions, the posterior distribution of A_1 shows some sensitivity—albeit minor—to the choice of its priors. These priors are centered at (in the case of A_2) and around (in the case of A_1) their true values of 0.6 and 0.8, respectively.

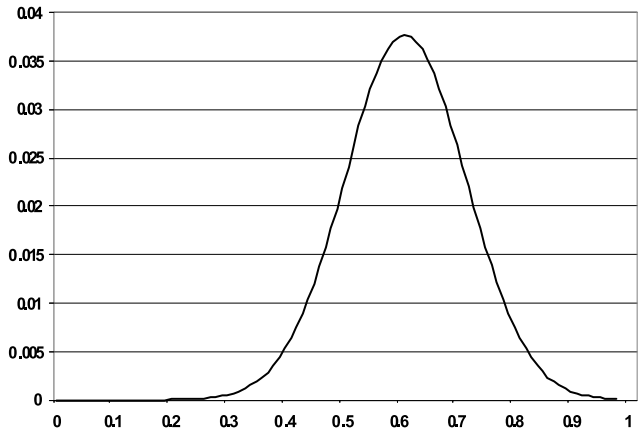


FIG. 9. *Posterior of* A_2 *based on* y_{12} *and* y_6.

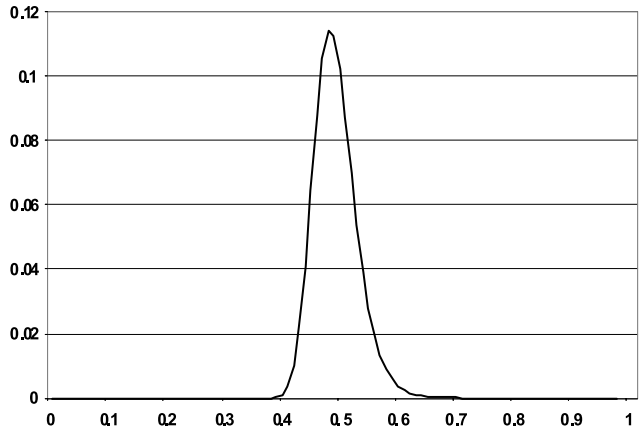


FIG. 10. *Posterior of* c *based on* y_{12} *and* y_6.

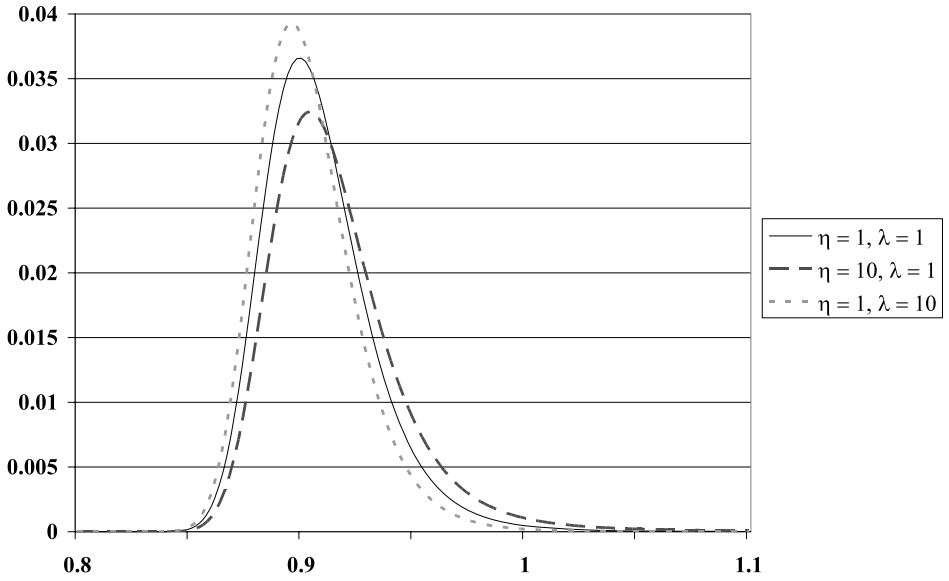


FIG. 11. *Posterior distribution of E_1 with different values of η and λ.*

Since the prior on c is a uniform on $(0, \omega/E_1)$, changing the value of ω would simply change the range of values that c can take. It will not change the shape of the posterior distribution of c. Finally, a use of the Pareto as a prior for E_1 results in a posterior distribution that looks much like that of Figure 7 produced by a gamma prior. This result—not illustrated here—is true irrespective of the choice of the hyperparameters of the Pareto prior. Indeed, the Pareto prior for E_1 indicates a higher degree of robustness of its resulting posterior as compared to the gamma prior.

Overall, it seems to be the case that the proposed procedure is robust to the choice of priors, and that the resulting posteriors converge to their correct values no matter the choice of priors.

5.2 Validation Against Physics Code Data

In this section we validate our approach using data pertaining to a pion that has been generated by a physics based code. These data are given in Table 2 and parallel those of Table 1, save for the fact that the data run from $t = 2$ to $t = 13$, and that $G(t)$, $t = 2, \ldots, 13$,

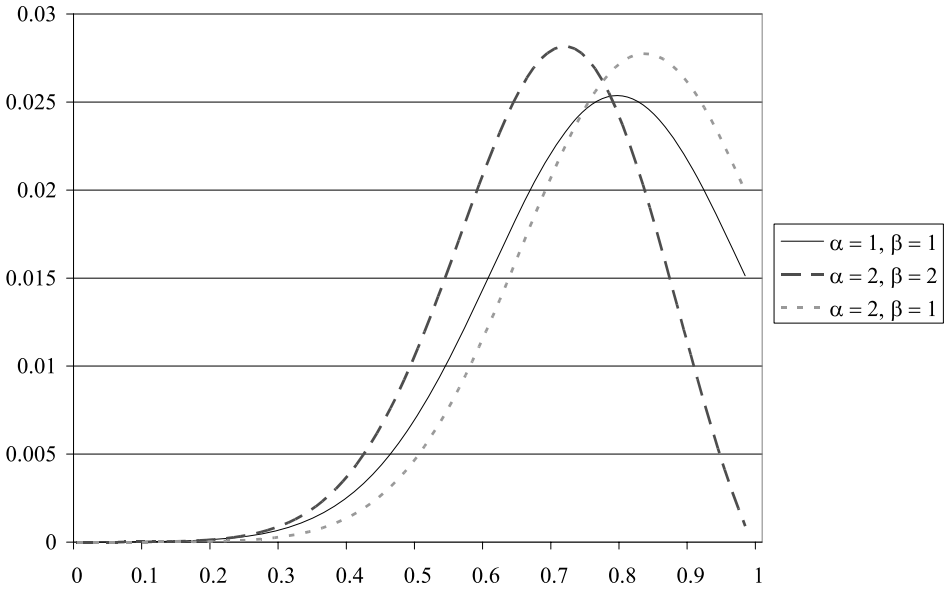


FIG. 12. *Posterior distribution of A_1 with different values of α and β.*

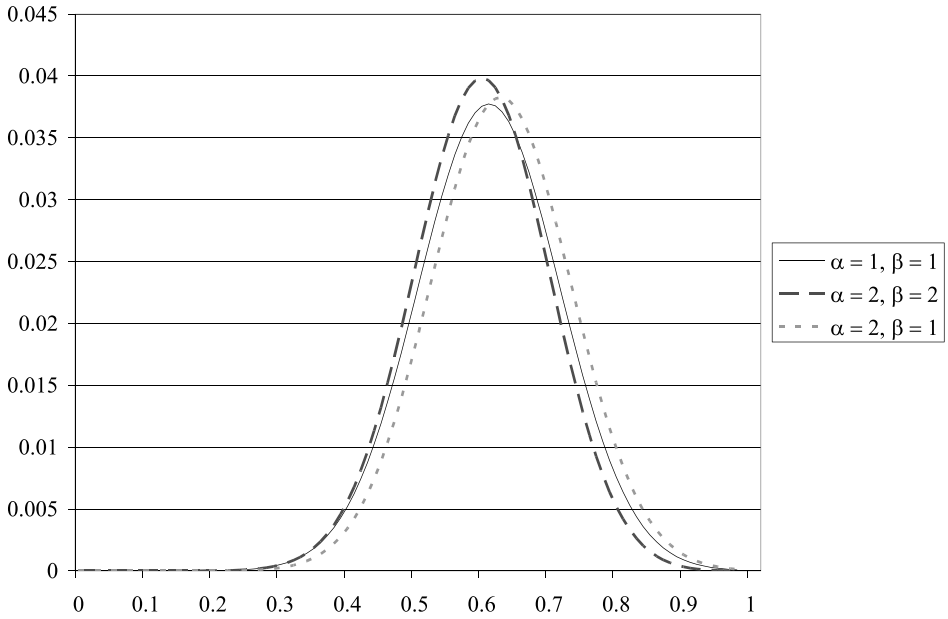


FIG. 13. *Posterior distribution of A_2 with different values of α and β.*

is not known. However, the Y_t's and their associated errors are provided by the code, the errors being a proxy for σ_t^2. The choice of t_1, t_2 and t_3 is based on the following consideration. By default, t_1 has to be the largest t for which the data are available; thus, in our case t_1 corresponds to $t = 13$. At t_1 all the exponential terms in equation (3.3) vanish. At t_2 we need to have the terms starting with e^{-2ct} vanish; this means that $t_2 \approx t_1/2$, which in our case would be 7. Similarly, $t_3 \approx t_1/3$, which is 4, and so on.

In Figure 14 (a), (b), (c) and (d), we show the posterior distributions of A_1, E_1, A_2 and c, respectively, based on y_{13} and y_7. The modes of these posterior distributions suggest the values of 0.52, 0.029, 0.02 and 0.4, for E_1, A_1, A_2 and c, respectively. The values of A_1 and E_1 given above are in good agreement with the values obtained by a physics based simulation code. Since the physics based codes are unable to obtain good estimates of A_2 and E_2 (equivalently, c), the results on A_2 and c obtained by us constitute a contribution toward the solution of an underlying scientific problem.

TABLE 2
Physics code based data on a pion

t	Index t_i	Y_t	Errors	
2		0.043865236	0.00013635	0.00014836
3		0.009347211	0.00008205	0.000089027
4	t_3	0.00406969	0.000051302	0.000057832
5		0.002187666	0.000031545	0.000034867
6		0.001252858	0.000018805	0.000018559
7	t_2	0.000735911	0.000011131	0.00001124
8		0.0004358	6.6393E−06	6.8252E−06
9		0.00025829	0.000004049	4.3108E−06
10		0.000153161	2.4808E−06	2.6302E−06
11		9.1412E−05	1.5264E−06	0.000001683
12		5.552E−05	9.5081E−07	1.0741E−06
13	t_1	3.54336E−05	6.3079E−07	7.0522E−07

Based on this exercise, plus others that are given in Landon (2007), our conclusion therefore is that the proposed approach is successfully validated against both simulated data as well as the physics code generated data. The exercises in Landon (2007) pertain to the quark masses of 4 photons and 5 pions.

6. EXTENDING THE APPROACH

The approach outlined in Sections 3 and 4 has some limitations. The purpose of this section is to prescribe strategies for overcoming these. By far, the most noteworthy limitation is that the model of equation (4.1) restricts attention to a consideration of the parameters A_1, E_1, A_2 and c, whereas the Lattice QCD equations have an infinite number of A_i's and E_i's. The second concern pertains to the fact that in Section 5, data associated with the t's intermediate to t_1 and t_2 are not used in the MCMC algorithm. The proposed approach therefore does not exploit all the available data y_t. Finally, there is a question of assuming a constant spacing c of

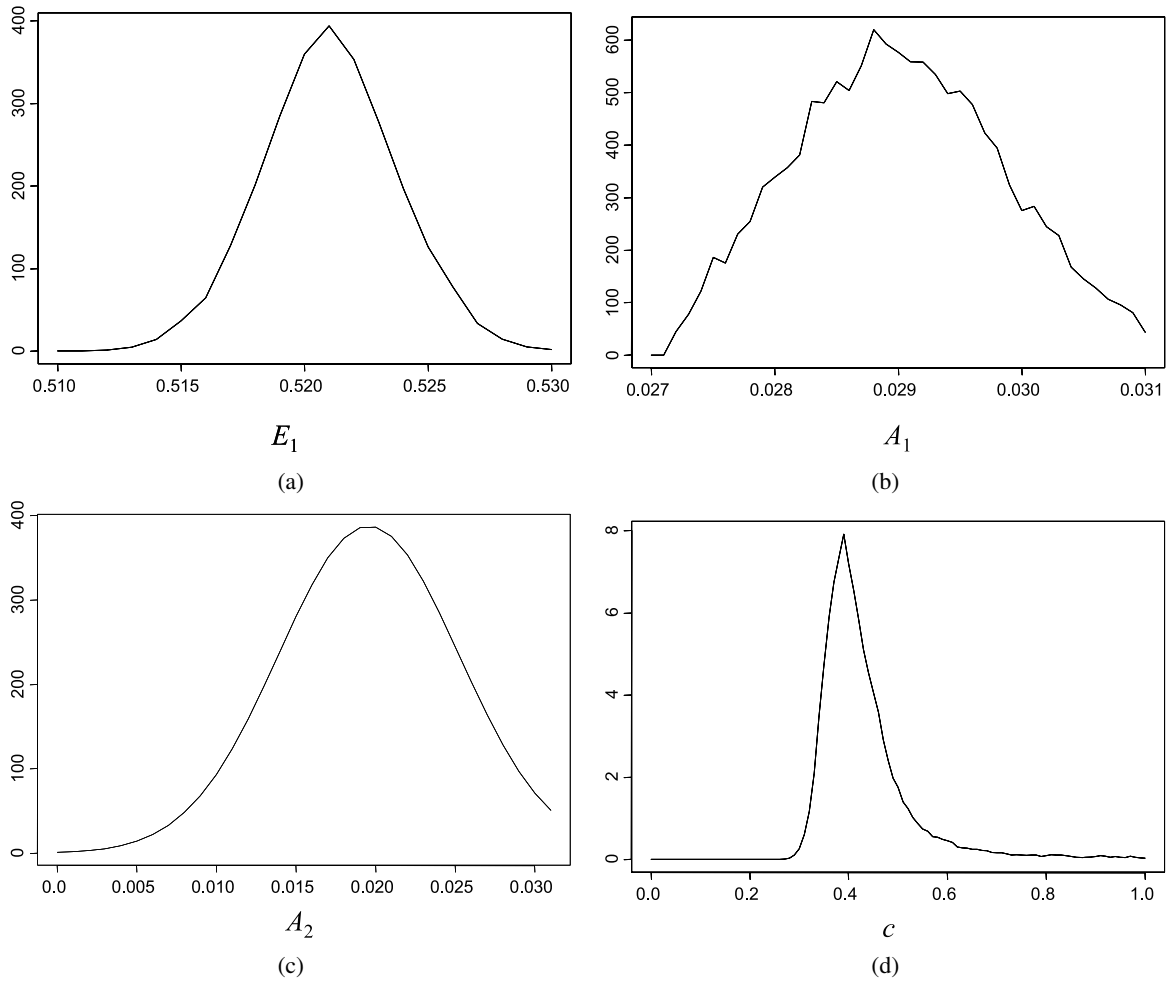


FIG. 14. *Posterior distributions of A_1, E_1, A_2 and c based on y_{13} and y_7.* (a) *Final posterior of E_1.* (b) *Final posterior of A_1.* (c) *Final posterior of A_2.* (d) *Final posterior of c.*

the E_i's. What is the effect of unequally spaced E_i's on inference? Recall that the role played by c is important. First, it imparts parsimony by eliminating all the E_i's save for E_1. Second, it gives birth to the telescopic series which is central to our approach. It turns out that the effect of c is transitionary (it is a nuisance parameter) and that inferences about E_2, A_3, E_3, $A_4, \ldots,$ are possible if we exploit a result observed in Section 5.

6.1 Inferences for E_2, A_3 and Beyond

Our ability to extend the approach of Sections 3 and 4 to the case of E_2, A_3, E_3, $A_4, \ldots,$ is driven by the feature noticed in Section 5.1, that y_6 does not contribute much toward the inferences for A_1 and E_1, beyond that provided by y_{12}. Thus, the effect of y_4, which corresponds to t_3 of Table 1, will be less so, making it possible for us to do the following:

Rewrite equation (3.1) as

$$G(t|\cdot) - A_1 e^{-E_1 t} = \sum_{n=2}^{\infty} A_n e^{-E_n t}, \tag{6.1}$$

and let $G^*(t|\cdot) = G(t|\cdot) - \widehat{A}_1 e^{-\widehat{E}_1 t}$, where $\widehat{A}_1$ and $\widehat{E}_1$ are the modes (means) of the posterior distributions of A_1 and E_1 obtained via the likes of Figures 7(b) and 8(b). Thus,

$$G^*(t|\cdot) \approx \sum_{n=2}^{\infty} A_n e^{-E_n t},$$

and setting $A_n^* = A_{n+1}$ and $E_n^* = E_{n+1}$, for $n = 2, 3, \ldots,$ we have

$$G^*(t|\cdot) \approx \sum_{n=1}^{\infty} A_n^* e^{-E_n^* t}. \tag{6.2}$$

The right-hand side of equation (6.2) parallels the right-hand side of equation (3.1), save for the fact

that A_n^* and E_n^* replace A_n and E_n. The material of Sections 3 and 4 now applies, but with the caveat that since equation (6.2) is an approximation, whereas equation (3.1) is exact, the variance of the error terms associated with the former should be larger than those associated with the latter.

The posterior distributions of E_1^* and A_2^* will be the posterior distributions of E_2 and A_3. The role of c as a nuisance parameter is now apparent. The posterior distribution of A_1^* will serve as a revised posterior distribution of A_2. Indeed, for the MCMC runs associated with the treatment of equation (6.2), we may sample from the posterior distribution of A_2 to generate the posterior distribution of A_1^*.

We may continue in the above vein to estimate E_3 and A_4 by defining $G^{**}(t|\cdot) = G^*(t|\cdot) - \widehat{A}_1^* e^{-\widehat{E}_1^* t}$, where $\widehat{A}_1^*$ and $\widehat{E}_1^*$ are the modes of the posterior distributions of A_1^* and E_1^*, respectively, and similarly with (E_4, A_5), (E_5, A_6), and so on.

6.2 Using Additional Y_t's

For enhanced inferences about the parameters A_1 and E_1 we may want to use all values of Y_t intermediate to those associated with the labels t_1 and t_2 of Tables 1 and 2 and, similarly, with the Y_t's intermediate to the ones associated with the labels t_2 and t_3, and so on. What makes this possible is the fact that t_1 is the *largest* value of t for which $(A_2, E_2), (A_3, E_3), \ldots,$ gets annihilated, whereas t_2 is the largest value of t at which $(A_3, E_3), (A_4, E_4), \ldots,$ gets annihilated, and so on. Thus, values of t intermediate to t_1 and t_2 will continue to annihilate $(A_2, E_2), (A_3, E_3), \ldots,$ and those intermediate to t_2 and t_3 will annihilate $(A_3, E_3), (A_4, E_4), \ldots,$ and so on.

Let $y_{11}, y_{12}, \ldots, y_{15}$ denote the Y_t's intermediate to those associated with the labels t_1 and t_2. Then, to incorporate the effect of $y_{11}, \ldots, y_{15}$ for enhanced inference about A_1 and E_1, the iterative scheme described in Phase I of Section 4.2 will have to be cycled five more times, each cycle involving a use of the y_{1j}, $j = 1, \ldots, 5$, before proceeding to Phase II, wherein the effect of y_6 (of Table 1) and y_7 (of Table 2) comes into play and, similarly, with y_{21}, the single value intermediate to that associated with the labels t_2 and t_3.

7. SUMMARY AND CONCLUSIONS

In this paper we have proposed and developed a statistical approach for addressing a much discussed problem in particle physics. Indeed, a problem that has spawned several Nobel prizes in Physics. The essence of the problem boils down to estimating a large (conceptually infinite) number of unknown parameters based on a finite number of nonlinear equations. Statisticians refer to such problems as large p—small n. Each equation in our problem comprises of the sum of several exponential functions.

Previous approaches for addressing this problem have been physics based—such as perturbation methods—and statistics based—such as chi-squared goodness of fit, and Empirical Bayes. Physicists have found such approaches unsatisfactory, and have called for a use of proper Bayesian approaches, thus this paper.

The Bayesian approach proposed by us has been facilitated by the fact that by introducing a latent parameter, the architecture of the nonlinear equations reveals an attractive pattern. This pattern boils down to our consideration of a truncated telescopic series of equations, each equation being the sum of a finite number of exponential functions. Similar sets of equations also arise in other arenas of science, as mentioned in Section 3.1. The nonlinear nature of the equations mandates that our proposed approach—which entails stylized proper priors—be implemented by a particle style Markov chain Monte Carlo (MCMC) approach. Such a procedure turns out to be computationally very intensive—about one million iterations for making inference about three parameters.

The proposed procedure, when invoked on simulated data, is able to reproduce the input parameters. This is one way to claim the validity of our approach. The procedure, when invoked on some real data pertaining to the quark masses of protons and pions, is also able to produce results that are in agreement with the results produced using alternate physics based methods. However, the physics based methods are able to obtain only partial results. By contrast, our approach can produce estimates of as many parameters as is desired—but there is no way to validate these against alternate approaches or actual numbers, because these are not available.

Future work in this arena will entail enhancements to gain computational efficiencies and the choice of proper priors that are motivated by a consideration of the underlying physics. This means that an undertaking such as this will call for insights and skills that go beyond mathematics, statistics and computing. Some appreciation of the underlying physics is necessary for, among other things, interest and inspiration! A referee of this paper has made the interesting suggestion of considering "reference priors." This we have been unable to do because, for the parameters in question, such priors are not readily available.

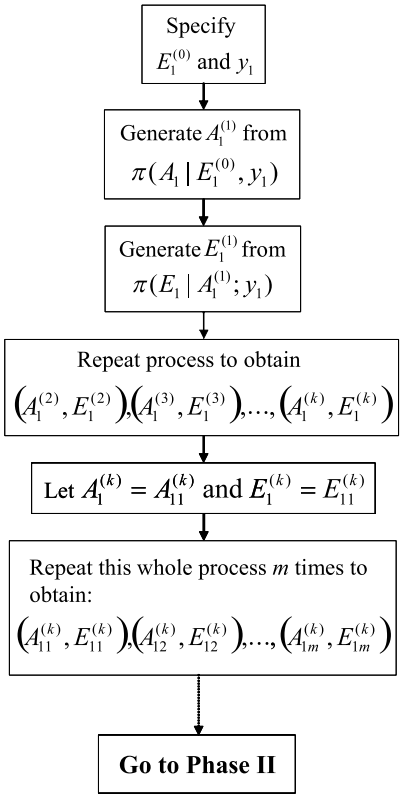
ACKNOWLEDGMENTS

The authors would like to thank Professor Ali Eskandarian of GWU for introducing us to this topic and for orchestrating our involvement in it. We want to acknowledge (the late) Professor Cornelius Bennhold, also of GWU, for contributing to our learning of the physics part of the problem, for keeping us honest and for providing us with the data to validate our approaches. The several helpful comments of the referees and the Editor, Professor David Madigan, are gratefully acknowledged. This research is supported by The Army Research Office Grant W911NF-09-1-0039 and by the National Science Foundation Grant DMS-09-15156 with The George Washington University. The work of Frank Lee is supported in part by U.S. Department of Energy under Grant DE-FG02-95ER-40907.

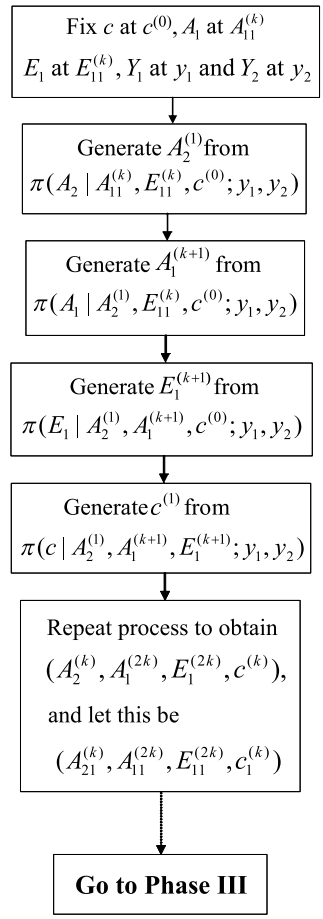
APPENDIX

Schemata of the 3 Phase MCMC algorithm, which can be downloaded at the following: http://www.gwu.edu/~stat/irra/Lattice_QCD.htm.

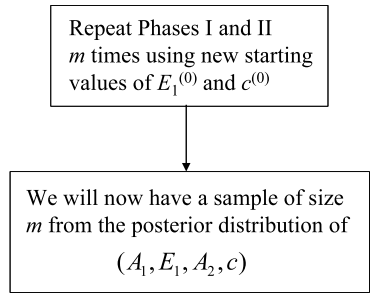
Phase I



Phase II



Phase III



REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York. MR0438625

BERNARDO, J.-M. and SMITH, A. F. M. (1994). *Bayesian Theory*. Wiley, Chichester. MR1274699

BRETTHORST, G. L., HUTTON, W. C., GARBOW, J. R. and ACKERMAN, J. J. H. (2005). Exponential parameter estimation (in NMR) using Bayesian probability theory. *Concepts in Magnetic Resonance Part A* **27A** 55–63.

CHEN, Y., DRAPER, T., DONG, S. J., HORVATH, I., LEE, F. X., LIU, K. F., MATHUR, N., SRINIVASAN, C., TAMHANKAR, S. and ZHANG, J. B. (2004). The sequential empirical Bayes method: An adaptive constrained-curve fitting algorithm for lattice QCD." *Phys. Rev. D*. Available at http://arxiv.org/pdf/hep-lat/0405001.

DYSON, R. D. and ISENBERG, I. (1971). Analysis of exponential curves by a method of moments, with special attention to sedimentation equilibrium and fluorescence decay. *Biochemistry* **10** 3233–3241.

DZIERBA, A., MEYER, C. and SWANSON, E. (2000). The search for QCD exotics. *American Scientist* **88**(5) 406–416.

FIEBIG, H. R. (2002). Spectral density analysis of time correlation functions in lattice QCD using the maximum entropy method. *Phys. Rev. D* **65** 094512.

FLEMING, G. T. (2005). What can lattice QCD theorists learn from nmr spectrocopists? Technical Report, Jefferson Labs., Newport News, VA.

FOSS, S. D. (1969). A method for obtaining initial estimates of the parameters in exponential curve fitting. *Biometrics* **25** 580–584.

GIURCĂNEANU, C. D., TĂBUŞ, I. and ASTOLA, J. (2005). Clustering time series gene expression data based on sum-of-exponentials fitting. *EURASIP J. Appl. Signal Process.* **8** 1159–1173. MR2168617

GLASS, H. I. and DE GARRETA, A. C. (1967). Quantitative analysis of exponential curve fitting for biological applications. *Physics in Medicine and Biology* **12** 379–388.

GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F* **140** 107–113.

GRIFFITHS, D. (1987). *Introduction to Elementary Particles*. Wiley, New York.

HILDEBRAND, F. B. (1956). *Introduction to Numerical Analysis*. McGraw-Hill, New York. MR0075670

LANDON, J. (2007). A problem in particle physics and its Bayesian analysis. Ph.D. thesis, George Washington Univ., Washington, DC. MR2710570

LEPAGE, G. P., CLARK, B., DAVIES, T. H., HORNBOSTEL, K., MACKENZIE, P. B., MORNINGSTAR, C. and TROTTIER, H. (2002). Constrained curve fitting. *Nuclear Physics B Proceedings Supplements* **106** 12–20.

MORNINGSTAR, C. (2002). Bayesian curve fitting for lattice gauge theorists. *Nuclear Physics B Proceedings Supplements* **109** 185–191.

NAKAHARA, Y., ASAKAWA, M. and HATSUDA, T. (1999). Hadronic spectral functions in lattice QCD. *Phys. Rev. D* **60** 091503.

PAGELS, H. R. (1982). *The Cosmic Code*: *Quantum Physics As the Language of Nature*. Simon and Schuster, New York.

PALUSZNY, M., MARTÍN-LANDROVE, M., FIGUEROA, G. and TORRES, W. (2008/09). Boosting the inverse interpolation problem by a sum of decaying exponentials using an algebraic approach. *Electron. Trans. Numer. Anal.* **34** 163–169. MR2597808

PERL, W. (1960). A method for curve-fitting by exponential functions. *The International Journal of Applied Radiation and Isotopes* **8** 211–222.

RIORDAN, M. and ZAJC, W. A. (2006). The first few microseconds. *Scientific American* **294** 34–41.

ROBERTSON, J. S. (1957). Theory and use of tracers in determining transfer rates in biological systems. *Physiological Reviews* **37** 133–157.

RUBINOW, S. I. (1975). *Introduction to Mathematical Biology*. Wiley, New York.

SANCHEZ-GASCA, J. J. and CHOW, J. H. (1999). Performance comparison of three identification methods for the analysis of electromagnetic oscillations. *IEEE Transactions on Power Systems* **14** 995–1002.

SINGPURWALLA, N. D. (2006). *Reliability and Risk*: *A Bayesian Perspective*. Wiley, Chichester. MR2265917

SMITH, R. E. and MORALES, M. F. (1944). On the theory of blood-tissue exchanges: II. Applications. *Bull. Math. Biol.* **6** 133–139.

VAN LIEW, H. D. (1967). Method of exponential peeling. *J. Theoret. Biol.* **16** 43–53.

WILCZEK, F. A. (2005). Asymptotic freedom: From paradox to paradigm. *Internat. J. Modern Phys. A* **20** 5753–5777. MR2189131

YAM, P. (1993). QED for QCD. *Scientific American* **269** 23–24.