SIAM Review
Vol. 55, No. 1, pp. 149–167

Estimating Parameters in Physical Models through Bayesian Inversion: A Complete Example*

Moritz Allmaras†
Wolfgang Bangerth‡
Jean Marie Linhart†
Javier Polanco†
Fang Wang†
Kainan Wang†
Jennifer Webster†
Sarah Zedler†

Abstract. All mathematical models of real-world phenomena contain parameters that need to be estimated from measurements, either for realistic predictions or simply to understand the characteristics of the model. Bayesian statistics provides a framework for parameter estimation in which uncertainties about models and measurements are translated into uncertainties in estimates of parameters. This paper provides a simple, step-by-step example—starting from a physical experiment and going through all of the mathematics—to explain the use of Bayesian techniques for estimating the coefficients of gravity and air friction in the equations describing a falling body. In the experiment we dropped an object from a known height and recorded the free fall using a video camera. The video recording was analyzed frame by frame to obtain the distance the body had fallen as a function of time, including measures of uncertainty in our data that we describe as probability densities. We explain the decisions behind the various choices of probability distributions and relate them to observed phenomena. Our measured data are then combined with a mathematical model of a falling body to obtain probability densities on the space of parameters we seek to estimate. We interpret these results and discuss sources of errors in our estimation procedure.

Key words. parameter estimation, Bayesian estimation techniques, priors, posterior probability distribution

AMS subject classifications. 34A55, 62F15, 62P35, 97M50

DOI. 10.1137/100788604

1. Introduction.

Much of the motivation of applied mathematics lies in our desire to understand the behavior of systems or to predict their behavior in various situations, for example, in order to optimize their design and improve performance.

*Received by the editors March 5, 2010; accepted for publication (in revised form) November 22, 2011; published electronically February 7, 2013.
http://www.siam.org/journals/sirev/55-1/78860.html

†Department of Mathematic Texas A&M University, College Station, TX 77843-3368. The work of the first, third, fifth, and eighth authors was supported by award KUS-C1-016-04 from the King Abdullah University of Science and Technology. The work of the seventh author was supported by U.S. Department of Homeland Security grant 2008-DN-077-ARI001-02.

‡Corresponding author. Department of Mathematics, Texas A&M University, College Station, TX 77843-3368 (bangerth@math.tamu.edu). The work of this author was supported by NSF award DMS-0604778, U.S. Department of Energy grant DE-FG07-07ID14767, U.S. Department of Homeland Security grant 2008-DN-077-ARI001-02, award KUS-C1-016-04 from the King Abdullah University of Science and Technology, and an Alfred P. Sloan Research Fellowship.

To this end, we come up with equations that describe physical, chemical, or biological processes. Solving these equations numerically with appropriate boundary conditions and forcing terms allows us to predict the system's response.

For example, we can model the flow of air around an airplane using the Navier–Stokes or Euler equations of gas dynamics to infer lift and drag forces; we can use the equations of elasticity to predict how the wings transfer lift forces to the fuselage; and we can use a multispecies, chemically reacting version of the Navier–Stokes equations to describe fuel combustion in the engines. However, in all of these cases, realistic predictions require that we also know the values of all material properties involved: density, viscosity, and thermal properties of air; density and elastic moduli of the materials of which the airplane is made; and chemical reaction rates of the components of fuel and air. In mathematics, we often take knowledge of these material properties for granted, but someone must have determined them once, and presumably some mathematics must have been involved in this process!

In the above examples, we wanted to know parameters so that we could predict the behavior of systems in response to external stimuli for which we do not want to perform actual experiments. But there are also cases where the parameters themselves are of interest. The prototypical case is imaging: If we could identify the distribution of materials inside the human body using data obtained from irradiating the body from different angles with X-rays, we could produce images of medical value; the fact that knowing this material distribution would also allow us to predict what would happen if we irradiated the body from another angle not yet tried is then of minor interest.

Whatever the reason, these examples show that we need to know the numerical values of parameters occurring in mathematical models. How to infer parameters from indirect measurements in physical systems is studied in the field of *parameter estimation*.[1] Parameter estimation uses a mathematical model of the system under consideration (e.g., the Navier–Stokes equations describing incompressible flow) that contains unknown parameters (e.g., the fluid's viscosity) as well as physical measurements (e.g., pressure or velocity data at individual points of the domain occupied by the fluid) to infer the numerical values of parameters that appear in the model.

The basic idea of parameter estimation is to find those parameters for which the predictions obtained from a model best match our measurements or observations in the actual physical system. The traditional, *deterministic* approach is to find *the best* set of parameters $\mathbf{m}^*$ among all parameters $\mathbf{m}$, i.e., those values for which the predicted behavior best matches actual measurements among all possible parameter values. There are different, competing ways to define what constitutes the best parameters. However, any approach that only produces a single best value $\mathbf{m}^*$ does not take into account the fact that our measurements contain measurement noise or that we might be unsure of the exact model, and that, consequently, we cannot be entirely sure of our estimated parameters. In other words, deterministic inversion methods do not explain the *uncertainty* in parameter estimates.

In this paper, we will consider an alternative, statistical approach to parameter estimation problems—called the *Bayesian* method—that takes a different viewpoint:

[1]Parameter estimation is the term commonly used for the discipline of determining material properties or other parameters in mathematical models from measured data. If these parameters appear in differential equations, in particular, partial differential equations, parameter estimation is often also referred to as an *inverse problem*. The process of solving a parameter estimation or inverse problem is called "inverting data for parameters."

Rather than trying to find *the* best parameter $\mathbf{m}^*$, it asks for a *probability density* $\sigma(\mathbf{m})$ for the parameters $\mathbf{m}$ that updates the prior information we have on $\mathbf{m}$, which is also given as a probability density, in light of the actual measurements that we will denote by $\mathbf{d}$. This probability density function (pdf) is computed taking into account the mathematical model, information about the measurement process, and the measured data, as well as any prior knowledge we may have about the parameters we seek; it is consequently a *posterior* probability density as we can only determine it *after* making actual measurements; it is conditioned on $\mathbf{d}$. We will henceforth write this density as $\sigma(\mathbf{m}|\mathbf{d})$.

The Bayesian solution of the problem is the entire posterior pdf, but in some cases we may choose to summarize some of the characteristics of this density. For example, we may compute the point(s) where the pdf achieves the maximum. Since such a maximizer is often used as an estimate of the parameters, it is called a "maximum a posteriori" or "MAP" estimator. In particular, if it is unique, then it is defined as

$$\mathbf{m}_{\mathrm{MAP}} = \arg \max_{\mathbf{m}} \ \sigma(\mathbf{m}|\mathbf{d}). \tag{1.1}$$

Note, however, that there need not be a single, unique point that maximizes σ. On the other hand, if the most likely values are all located in one region of parameter space, it may make more sense to compute the "posterior mean"

$$\mathbb{E}(\mathbf{m}|\mathbf{d}) = \int \mathbf{m}\ \sigma(\mathbf{m}|\mathbf{d})\ \mathrm{d}\mathbf{m}, \tag{1.2}$$

which provides some information on where the pdf is centered. Information about the scatter around the posterior mean of the ith component of $\mathbf{m}$ is provided by the posterior standard deviation defined as

$$\begin{aligned} \mathrm{stddev}(m_i|\mathbf{d}) &= \sqrt{\mathrm{var}(m_i|\mathbf{d})} \\ &= \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |m_i - \mathbb{E}(\mathbf{m}_i|\mathbf{d})|^2\ \sigma(\mathbf{m}|\mathbf{d})\ \mathrm{d}m_1\ \mathrm{d}m_2}. \end{aligned} \tag{1.3}$$

It is clear that $\sigma(\mathbf{m}|\mathbf{d})$ contains more information than just the posterior means and standard deviations. For example, if $\sigma(\mathbf{m}|\mathbf{d})$ is bi- or multimodal, i.e., there is more than one region of parameter space in which $\sigma(\mathbf{m}|\mathbf{d})$ has local maxima (indicating that parameters there are likely because they match measurements $\mathbf{d}$ well), then $\mathbf{m}_{\mathrm{MAP}}$ will only reveal the region in which probabilities are, in fact, largest, and $\mathbb{E}(\mathbf{m}|\mathbf{d})$ may lie in a region where parameters are altogether unlikely. Knowing the actual probability density σ can help avoid both of these pitfalls.

Objective and Outline. In this paper, we will consider a simple model problem in parameter estimation and show how the posterior probability density $\sigma(\mathbf{m}|\mathbf{d})$ can be computed. Our aim is to illustrate the techniques of Bayesian parameter estimation as well as to build an intuitive understanding of what it means to solve the problem of "inverting" measurements for parameters. To this end, we will first present a mathematical model describing a body falling in air in section 2. Section 3 then discusses what we know about the parameters of this model (section 3.1), the measurements we have obtained in our experiments along with their uncertainty (section 3.2), and how model and data can be combined to estimate parameters (section 3.3). We present a first set of results in section 4. Upon closer examination, these estimates turn out to miss an important aspect of our experiment, and we briefly mention techniques

for validating the correctness of estimates in section 5 before discussing an approach to correcting the initial oversight in section 6. We draw conclusions and present an outlook in section 7.

Background Notes. This study grew out of our weekly seminar in which we read and discussed Albert Tarantola's book on inverse problem theory [24], a book that has deeply influenced many applications scientists—particularly in the geosciences—who first came to know about Bayesian inversion through the ideas laid out in it. While many of those ideas may have been around when its first edition was published in 1987 [23], it appears to have been the first time a comprehensive framework of the statistical (Bayesian) viewpoint of inverse problems was made accessible to scientists in the applied sciences. Newer books in this area that stay somewhat closer to the usual statistical terminology and methodology have appeared since then, such as the one by Kaipio and Somersalo [14]. Nevertheless, Tarantola's book has remained an oft-cited and hugely influential book and has become widely available through SIAM's inexpensive, significantly revised edition of 2004 [24].

Tarantola's book provides a theoretical background for further discussion of parameter estimation problems. However, it lacks simple, concrete examples to which its techniques can be applied and, as a consequence, is sometimes a bit too abstract as an introductory text for students; this paper strives to close this gap. Tarantola's book is also written using nonstandard statistical terminology and notation that make it difficult for students to relate it to other textbooks. In this note we will, therefore, stay closer to the usual terminology.

From a practical perspective, the Bayesian inversion techniques discussed here are relatively straightforward to translate to actual applications only if the number of parameters is small, say, two or three. The reason is that one eventually needs to form integrals over the space of parameters, and high-dimensional integrals cannot be performed with reasonable computational effort through simple quadrature methods that are available to most advanced undergraduates. The application we consider here falls into the category of simple-enough problems, and we will therefore not have to deal with this complexity. However, most practical cases have many parameters, and integrals will then have to be approximated in some other way. Ultimately, Bayesian inversion techniques have only become successful over the past two or three decades because of the concurrent development of a technique that provides exactly this high-dimensional integration: Markov chain Monte Carlo (MCMC) sampling methods [7,8, 14]. Bayesian methods, therefore, almost always rely on MCMC sampling in practical applications. However, since the purpose of this paper is not the introduction of MCMC methods but an expository presentation of the ideas of Bayesian inversion using a concrete example, we will forgo the discussion of these topics and only refer to them in places where they would be necessary for more complex problems.

2. The Model Problem: A Body Falling in Air. For this study, we looked for a problem that is simple to understand and for which measurements can easily be obtained. We chose to investigate whether we can determine the gravitational acceleration and air friction coefficients by measuring the distance, $z(t)$, that a body dropped from a certain height has fallen at time t. This section describes the physical model, while the next section will deal with obtaining measurement data.

A body dropped from a height follows Newton's second law of motion,

$$F_{\text{total}} = m\frac{dv}{dt}, \tag{2.1}$$

where F_{total} is the sum of forces acting on the body, m is its mass, and $v(t) = dz/dt$ is its velocity. We will measure distance $z(t)$ and velocity $v(t)$ in the downward direction.

What are the forces acting on the falling body? First, there is gravity, which can be expressed by mg, where g is the acceleration due to gravity and is one of the two parameters that we want to determine. Second, the body is decelerated due to air resistance, which is a frictional force on the body.

Let us discuss this decelerating force further. Air is a fluid, and when an object falls through air fast enough, there is a turbulent boundary layer behind the object. In that case, the air resistance is proportional to the square of its velocity, v. The constant of proportionality, k, is called the coefficient of air resistance. k can be computed from a number of more fundamental parameters including the density and viscosity of air and the cross-section of the object. For simplicity we will not attempt to identify these independently, though procedures for the task of determining high Reynolds number drag coefficients are available in the literature (see, for example, [21, Chapter 5] and [16, pp. 46–54]). We are also not interested in the mass of the body, so let us denote by $C = k/m$ the *specific* coefficient of air resistance. Let us assume that g and C are the two parameters we would like to determine from experiments.

The considerations above lead us to conclude that the total force that acts on the object is

$$F_{\text{total}} = mg - mCv^2. \tag{2.2}$$

We combine Newton's second law of motion (2.1) and the sum of forces (2.2) to obtain the following set of differential equations describing the motion:

$$\frac{dv}{dt} = g - Cv^2, \qquad \frac{dz}{dt} = v. \tag{2.3}$$

Initial conditions are obtained by assuming that the body is released at time t_0 and that it is at rest initially:

$$z(t_0) = 0, \qquad v(t_0) = 0. \tag{2.4}$$

These equations can be integrated to obtain an analytical solution:

$$v(t) = v_\infty \tanh\left[\frac{g(t-t_0)}{v_\infty}\right], \qquad z(t) = \frac{1}{C}\log\cosh\left[\frac{g(t-t_0)}{v_\infty}\right]. \tag{2.5}$$

Here, we have introduced a "terminal" velocity $v_\infty = \sqrt{mg/k} = \sqrt{g/C}$. This velocity can be interpreted by considering that we expect the velocity of a falling body to continuously increase, but that asymptotically air resistance will balance gravity and the velocity will approach a constant terminal velocity. The expression for v_∞ then follows from the balance equation $F_{\text{total}} = 0$.

We want to estimate g and C. Both of these could be determined individually through separate experiments. In fact, this is what is commonly done. Gravity can easily be determined by measuring the gravitational force ("weight") on objects of known mass [27], or by measuring the period of a pendulum [11, Chapter 4.1].[2] For

[2]The local gravity constant can also be approximated to very high accuracy using global models by expressing gravity on Earth's surface in terms of spherical harmonics. Such models were originally obtained by fitting known measurements [12,13]. They were later computed by analyzing the motion of satellites, dating back to the very first objects in orbit in 1958 [15], and are today determined to such an accuracy that changes on the order of a few centimeters in the ground water level can be detected in the gravity signal [10]. See [2] and the references cited therein for an overview of current satellite geodesy.

finding drag coefficients, see [16,21]. Individual experiments are not always possible in more complicated systems, and the purpose of this paper is to show how parameters and their probability distributions can be determined from simple experiments that are not tailored to the estimation of a single parameter.

As a final note in this section, let us remark that the model used here admits an analytic solution, given in (2.5). For what follows this is not necessary at all: all we need to be able to do is to compute $z(t_i)$ at times t_i for given parameters g and C. An analytic solution for this is fine, but if, as is common for most mathematical models, an analytic solution cannot be found, then we can simply integrate the differential equation numerically to obtain an approximate value computationally (for the numerical solution of ordinary differential equations, see, for example, [17, Chapter 8]).

3. Statistical Description of the Inverse Problem. Our goal is to derive a *posterior* probability density $\sigma(\mathbf{m}|\mathbf{d})$ that, speaking in layman's terms, describes the updated probability of the parameter $\mathbf{m}$ in light of the measurement data $\mathbf{d}$.[3] Following the Bayesian approach to inverse problems, we need to define the several quantities that enter the formulas we will use to compute σ:

- The *prior probability*: We will define a probability density $\rho_M(\mathbf{m})$ through which we encode prior knowledge about the parameters, such as the positivity of g.
- The *likelihood*: Given a particular parameter $\mathbf{m} = \{g, C\}$, we can predict how the system will behave. Let us define the vector $\mathbf{f}(\mathbf{m})$ with entries $f_i(\mathbf{m}) = z(t_i)$, where $z(t_i)$ is the distance the body has fallen at time t_i as given by (2.5), for the given set of parameters $\mathbf{m}$. In practice, we will not be able to measure these particular values, but rather data of the form $\mathbf{d} = \mathbf{f}(\mathbf{m}) + \varepsilon$, where ε encodes measurement errors that introduce random uncertainty in the data. The random behavior of $\mathbf{d}$ for the fixed parameters controlling the falling body is described by the probability density $\rho_D(\mathbf{d}|\mathbf{m})$. It is called a likelihood when considered as a function of $\mathbf{m}$ with fixed $\mathbf{d}$. Note that $\rho_D(\mathbf{d}|\mathbf{m})$ is simply the pdf of the noise ε shifted by $f(\mathbf{m})$.
  We will derive a particular form of $\rho_D(\mathbf{d}|\mathbf{m})$ as a function of other parameters that control the variability of the measurements. We could make these parameters random variables that need to be estimated as well; however, pragmatically, we will obtain them by simply looking closely at the individual frames of our video showing the body falling.

An intuitive interpretation of how Bayesian inversion uses these two pieces of information to compute $\sigma(\mathbf{m}|\mathbf{d})$ is that we start with our prior knowledge (encoded in the prior probability density for parameters) and update it using the measurement data we have collected and the associated likelihood. We will discuss these two pieces in turn in sections 3.1 and 3.2, and then derive the final form of the posterior probability density in 3.3.

3.1. Using Prior Knowledge. The prior information about the parameters of the inverse problem is, in principle, independent of the measured data. In the Bayesian approach this prior knowledge is encoded in a probability distribution on the space

[3] This statement uses layman's language because the probability that $\mathbf{m}$ takes any particular value is zero if the posterior distribution of $\mathbf{m}$ has a pdf. Rather, a more correct, though less intuitive, statement would be that $\sigma(\mathbf{m}|\mathbf{d})$ describes the updated probability that the parameter takes values in an arbitrary (measurable) subset A of parameter space given the data; it is given by $\int_A \sigma(\mathbf{m}|\mathbf{d})\, \mathrm{d}m$.

of parameters, whose density we denote by $\rho_M(\mathbf{m})$, where the index M indicates the "model" (a term often used in this context for "parameters" and not to be confused with the mathematical forward model (2.5)). In practice, one often uses *empirical Bayesian methods* where the data are also used to estimate constants that parametrize the prior distribution [5].

To define our prior knowledge about possible values of the parameters g and C, note that gravity points down and therefore must, given our choice of coordinate system, be positive. Likewise, it would be unphysical if the air friction coefficient C were negative. Furthermore, let us assume that a simple, preliminary experiment has convinced us that $g \leq 20\ \mathrm{m/s^2}$ and that $C \leq 0.5\ \mathrm{m^{-1}}$. Together, we can encode this information by assuming that g and C are independent random variables, g uniformly distributed on $\left[0, 20\mathrm{m/s^2}\right]$ and C uniformly distributed on $\left[0, 0.5\mathrm{m^{-1}}\right]$. That is, we use the following uniform *prior probability density* for the parameters:

$$\rho_M(\mathbf{m}) = \rho_M(g, C) = \kappa_M\ \chi_{[0,20\mathrm{m/s^2}]}(g)\ \chi_{[0,0.5\mathrm{m^{-1}}]}(C), \tag{3.1}$$

where κ_M is a normalization constant and χ_A is the characteristic function of the set A. This definition says that we know with certainty that $0\,\mathrm{m/s^2} \leq g \leq 20\,\mathrm{m/s^2}$ (i.e., the probability for g being outside this interval is zero), and we assume (probably because we do not know anything that indicates the opposite) that all values in this interval are equally likely; similarly for C.

Clearly, if our initial experiments gave us more information, either about the possible range of parameters or about areas within this range where they may be more plausibly located, we could incorporate this into $\rho_M(\mathbf{m})$. In either case, this is the best we can do for now, but we will use the measurement data discussed in the next section to *update* this prior probability to obtain a better description of which parts of parameter space are more or less likely, and we will show our results in section 4.

3.2. The Measurement Data and Their Uncertainties. We obtained actual measurement data at certain times t_i by analyzing frames of a video that records an object—a brown paper bag weighed down from the inside by a stuffed animal called *Mr. Tiger*[4]—falling along a tape measure (see Figure 3.1). The camera records 30 frames per second [4], so if we fix t_0 to be the initial time and $i = 0$ to be the number of the first video frame in which the body is seen to be falling, then $t_i = t_0 + i/30$. We visually inspect each frame and record the distance d_i that the body has fallen in frame i.

In an ideal experiment, d_i would equal the solution $z(t_i)$ for the "correct" values of gravity and air friction, but in practice we have measurement errors. These include

- the low resolution of our video camera, which blurs the markings on the tape that indicate full inches and feet of distance;
- rotation of the object as it is falling, making it difficult to identify the outline of the object and to define how far the body as a whole has traveled;
- the blurring of the object on each individual frame due to its own speed.

These sources of error are illustrated in Figure 3.2 (in fact, there is another somewhat hidden source of error, to which we will return in section 6). As a consequence, what we obtain from each frame of the camera is an estimate d_i of the actual distance, and a measure of uncertainty δ_i in d_i. We identify d_i and δ_i with the midpoint and half-width of the interval delineated by two white lines in Figure 3.2. The data $\mathbf{d} = (d_i)_{i=0}^N$ and the uncertainty parameter $\boldsymbol{\delta} = (\delta_i)_{i=0}^N$ are shown in Figure 3.3.

[4]No animals, real or imaginary, were harmed in the performance of these experiments.

Fig. 3.1 *Snapshot of the experimental setup from a distance, showing an object during its fall and the tape measure used for measurements. The image has been extracted from frame-by-frame video footage.*

Fig. 3.2 *Major sources of uncertainty in reading data. Left: blurred black marks on the tape measure indicating full feet (30.48 cm) of distance. Right: the error range in determining distances an object has fallen.*

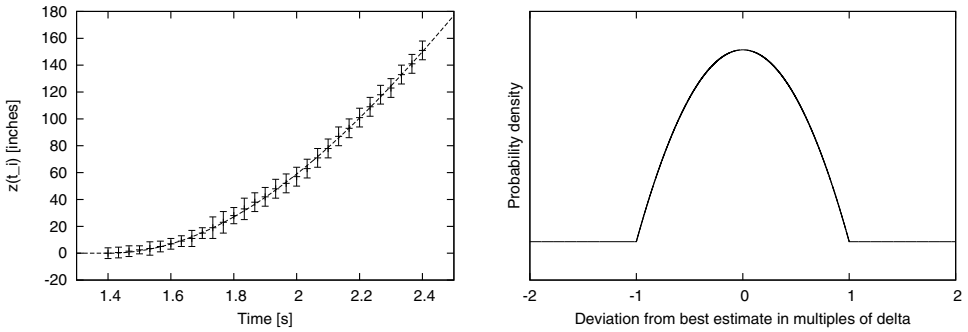


Fig. 3.3 *Left: raw data showing the distance the body has fallen at a given time after the camera was started. The household tape measure we used was purchased in the United States and therefore shows distances in inches and feet, which we reproduce here. All data were subsequently converted into the international unit system. The continuous line shows the predicted fall distance with the parameter estimates from section* 4*. It is obvious that we have been pessimistic about the magnitude of the error in most cases. Right: probability density* $\rho_i(d_i|\mathbf{m})$ *used for all data points.*

Let us again phrase this measurement process in the language of probability densities. We introduce a probability density $\rho_i(d_i|\mathbf{m})$ that describes the relative *likelihood* that the predicted fall distance $f_i(\mathbf{m})$ is compatible with our measurement d_i, given the fixed parameter $\mathbf{m}$. Note that the density $\rho_i(d_i|\mathbf{m})$ depends on the uncertainty parameter $\boldsymbol{\delta}$, so it would be more appropriate to write $\rho_i(d_i|\mathbf{m}, \boldsymbol{\delta})$.

Looking at the figure, we realize that we should choose a probability density that is centered around some "best guess value," and we will allow the probability to be nonzero around this value only within a range $\pm\delta_i$, as we are certain that a parameter that predicts a fall distance outside this range is incompatible with our measurements; for example, in Figure 3.2(b), the probability density $\rho_i(d_i|\mathbf{m})$ will be zero outside the range indicated by the two white lines as we can emphatically state that the box has fallen no less than the upper line and no more than the lower one, even if we may not be quite so sure about possible values between these lines.

To give a concrete form for $\rho_i(d_i|\mathbf{m})$ is more difficult, as this qualitative discussion does not uniquely define a density. We could say that we do not really know what the correct value within this range is and assume that d_i is uniformly distributed on the interval $[f_i(\mathbf{m}) - \delta_i, f_i(\mathbf{m}) + \delta_i]$. On the other hand, we do believe that the true value was closer to the center of the interval than to the end points. We can express this using a probability density based on an inverted parabola:

$$\rho_i(d_i|\mathbf{m}) = K_i \left(1 - \frac{|d_i - f_i(\mathbf{m})|^2}{\delta_i^2}\right) \chi_{[f_i(\mathbf{m})-\delta_i, f_i(\mathbf{m})+\delta_i]}(d_i). \tag{3.2}$$

With this particular choice, the normalization constant is $K_i = 3/(4\delta_i)$, and the probability density is shown in the right panel of Figure 3.3. This choice of $\rho_i(d_i|\mathbf{m})$ is somewhat arbitrary, albeit consistent with our knowledge of uncertainties in the data. Choosing a different description of $\rho_i(d_i|\mathbf{m})$—for example, the triangular "hat function," a truncated Gaussian, or a beta distribution—will lead to different reconstructed values for g and C, though we have found that the dependence on the particular shape of $\rho_i(d_i|\mathbf{m})$ is not very strong.

Because we extract measurements from individual frames, it is reasonable to assume that the measurement errors are statistically independent. The probability density that describes the collection of all measurements is then a product of the densities for each measurement:

$$\rho_D(\mathbf{d}|\mathbf{m}) = \prod_i \rho_i(d_i|\mathbf{m}). \tag{3.3}$$

3.3. Inverting Data for Parameters. The approach to parameter estimation in Bayesian inversion is to compute—using our mathematical model based on the physics, the likelihood function, the prior distribution, and the measured data—a probability density $\sigma(\mathbf{m}|\mathbf{d}) = \sigma(g, C|\mathbf{d})$ that will allow us to answer questions like, "*What is the probability that g is between a and b, and C between d and e?*" If we can compute such a probability density, the answer to the question is $\int_a^b \int_d^e \sigma(g, C|\mathbf{d})\, \mathrm{d}C\, \mathrm{d}g$. Note that the concept of a probability density σ goes beyond the goal of just finding an *estimate* of the parameters $\mathbf{m} = \{g, C\}$. We can still define estimators based on the posterior density (e.g., a MAP estimator or the posterior mean), but clearly σ contains more information, while the MAP and the posterior mean are just two characteristics of the posterior.

In a proper application of Bayesian methods to inversion, the prior distribution should not swamp the data and should introduce as little *artificial* information as possible. The degree to which the data and the prior constrain the parameters is then reflected in the shape of the posterior pdf σ. In the ideal case, σ is well localized and unimodal, which would put a tight control on the variability of the parameters. If σ is unimodal but has a high variance, then the data may be too noisy and/or the prior might not provide enough information to constrain the parameters. A multimodal posterior may be an indication that more information is required to properly identify the parameters.

Given these considerations, let us return to the task of *computing* $\sigma(\mathbf{m}|\mathbf{d})$. Recall that if we knew the "exact" parameter $\mathbf{m}$ and if our model was correct, then we would measure $\mathbf{d} = \mathbf{f}(\mathbf{m})$ if there was no measurement noise. On the other hand, given measurement uncertainties, the likelihood of measuring $\mathbf{d}$ is given by $\rho_D(\mathbf{d}|\mathbf{m})$ as specified by (3.2) and (3.3).

However, we do not know $\mathbf{m}$. Rather than predicting $\mathbf{d}$ from a known $\mathbf{m}$ (the forward problem), we want to estimate $\mathbf{m}$ from measured values of $\mathbf{d}$ (the inverse problem). In other words, rather than computing a probability density for $\mathbf{d}$ given $\mathbf{m}$, we want a probability density for $\mathbf{m}$ given $\mathbf{d}$, taking into account our prior probability density $\rho_M(\mathbf{m})$. The tool for this is Bayes' theorem [14, 24], which allows us to derive the second probability density from the first and which in the current context reads

$$\sigma(\mathbf{m}|\mathbf{d}) = k\, \rho_M(\mathbf{m})\, \rho_D(\mathbf{d}|\mathbf{m}), \tag{3.4}$$

where k is a normalization constant that may depend on $\mathbf{d}$.

For our two-dimensional problem, it is easy to visualize the basic shape of the posterior density; we do not even need to compute k. We start by selecting a grid of points for $\mathbf{m}$, and for each value of $\mathbf{m}$ we compute the value of the prior density $\rho_M(\mathbf{m})$. Then we calculate the predicted measurements $\mathbf{f}(\mathbf{m})$ using the forward model and evaluate the corresponding likelihood $\rho_D(\mathbf{d}|\mathbf{m})$ that relates our predictions to the actual measurement values. We can also find the MAP by numerically approximating the integral $k^{-1} = \int \rho_M(\mathbf{m})\rho_D(\mathbf{d}|\mathbf{m})\, \mathrm{d}m$, and we can then use (1.2) and (1.3) to calculate the posterior mean and standard deviation by numerical integration.

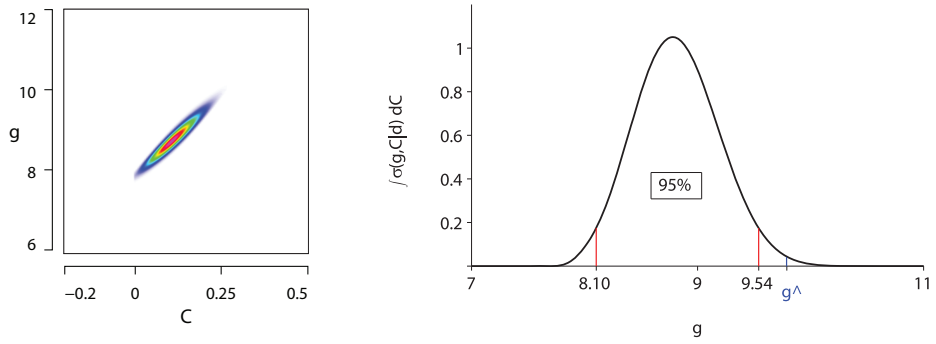


Fig. 4.1 *Left: a visualization of the posterior probability density $\sigma(g, C|\mathbf{d})$. Areas of highest probability are shown in red. Areas shown in white have a zero or negligible probability density. The color scale is in arbitrary units since we work with a nonnormalized probability density. Right: marginal posterior density for gravity $\sigma(g|\mathbf{d})$. Indicated in red is the smallest interval of 95% credibility; the true value of gravity $\hat{g} = 9.7935 m/s^2$ is shown in blue.*

We end this section by noting that more complex, high-dimensional problems cannot usually be solved by numerical quadrature to approximate integrals such as those necessary to compute k or to evaluate (1.2) and (1.3): if we covered parameter space with a grid for quadrature that has N points in the direction of each of n parameters, we would need to evaluate $\sigma(\mathbf{m}|\mathbf{d})$ at N^n points. This quickly becomes impractical unless n is very small—an effect called the *curse of dimensionality*. Rather, one usually has to resort to the heavier machinery of statistical sampling to approximate integrals, for example, using MCMC methods [7, 8, 14].

4. First Results. Let us finally bring theory and experimentation together. We are interested in the posterior probability density $\sigma(\mathbf{m}|\mathbf{d}) = \sigma(g, C|\mathbf{d})$ from (3.4), using the prior information defined in (3.1), which contains our expectations of reasonable physical values of the parameters, and the information from our measurements and their uncertainties, defined in (3.2) and (3.3).

As explained above, for a problem with only two parameters—a low-dimensional problem by most measures—a simple way to visualize $\sigma(g, C|\mathbf{d})$ is to evaluate it at a number of sample points $g_0 + i\Delta g, C_0 + j\Delta C$ and plot a function that interpolates the values at these points. Such a plot of the nonnormalized probability density for different values of the physical parameters $\mathbf{m} = \{g, C\}$ is shown in the left panel of Figure 4.1. As can be seen from the elongation of the region with high probability density, the parameters we identify are correlated. We will comment again on the cause of correlation in section 7.

Using formula (1.2), we compute the expected values for the parameters by approximating the integrals with appropriately weighted sums (quadrature) over the same set of mesh points from which the figure was generated:

$$\mathbb{E}(g|\mathbf{d}) = 8.82\frac{\mathrm{m}}{\mathrm{s}^2}, \qquad \mathbb{E}(C|\mathbf{d}) = 0.116\mathrm{m}^{-1}. \tag{4.1}$$

Here, we have used a mesh with 401 points in the range $7\mathrm{m/s}^2 \le g \le 11\mathrm{m/s}^2$ and 351 points for $0 \le C \le 0.35\mathrm{m}^{-1}$. We have verified that this is fine enough to make the quadrature error negligible. The posterior probability density σ is negligible outside these ranges for g, C.

While we have no comparison for C (since it depends on the viscosity and density of air as well as weight and cross-section of the dropped object), the value for gravity above is not a bad estimate for an experiment as simple as ours: at the location of College Station, TX ($\phi = 30.61°$N, $\psi = 96.32°$W, and $h = 298$ ft above sea level), an approximate value of true gravity can be computed as [26]

$$\hat{g} = \Big[9.780327\left(1 + 0.0053024 \cdot (\sin\phi)^2 - 0.0000058 \cdot (\sin 2\phi)^2\right) - 1.965 \cdot 10^{-6}\mathrm{m}^{-1} \cdot h\Big] \frac{\mathrm{m}}{\mathrm{s}^2} \approx 9.7935 \frac{\mathrm{m}}{\mathrm{s}^2}. \tag{4.2}$$

In other words, if we use the posterior mean as an estimator of the true value of gravity, we see that it underestimates the true g by some 10%. But the Bayesian solution is the posterior pdf, which contains a lot more information than just the expected values (4.1). We can, for example, look at the probability density for the gravity constant alone—the marginal posterior density—which is given by

$$\sigma(g|\mathbf{d}) = \int_{-\infty}^{\infty} \sigma(g, C|\mathbf{d})\, \mathrm{d}C.$$

Again, we approximate the integral by weighted sums over our sample points. The resulting marginal posterior density is shown in the right panel of Figure 4.1. The true value of gravity lies outside the interval $[8.10\mathrm{m/s}^2, 9.54\mathrm{m/s}^2]$ of 95% credibility (indicated in red). This implies that in terms of the posterior density we computed from our experiment, the value $\hat{g}$ (that we happen to know is correct) is a highly unlikely value for the gravity constant as determined from our experiment. Put another way, given that we know the correct answer, the results we get from using our prior knowledge and measurement data are not actually all that impressive! We explore one possible explanation for this discrepancy in section 6.

5. Validating Results. The result of our computations above is the probability density $\sigma(\mathbf{m}|\mathbf{d})$ derived in (3.4) from the measured data $\mathbf{d}$, prior probability ρ_M, and our assumptions on measurement errors ρ_D and the physical model (2.5). However, it is important to realize that we would have obtained such a probability *even if some of our assumptions had been completely unrealistic*. For example, if we had assumed that the physical model yields a linear fall distance $z(t) = at + b$, we could have made $\mathbf{m} = \{a, b\}$ the parameters to be identified and would have obtained a posterior probability density $\sigma(\{a, b\}|\mathbf{d})$ that would have indicated which values for a, b are the most probable. Of course, looking at where data points lie in Figure 3.3, it is clear that even the most probable linear model will describe the data in an entirely inadequate way!

It is therefore important to *validate* the results we have obtained above: Rather than simply trusting our results, we have to verify that they make sense. There are many ways to define measures that quantify whether the parameters identified by $\sigma(\mathbf{m}|\mathbf{d})$ reasonably describe the data $\mathbf{d}$ and their uncertainty (see, for example, [25]). Since most of these methods require sampling from the probability density $\sigma(\mathbf{m}|\mathbf{d})$—and therefore the machinery of MCMC methods [7, 8, 14] that are outside the scope of this paper—we will not describe them in detail here. That said, the general idea of such validation attempts is to generate a collection of *synthetic measurements* $\mathbf{d}^{(k)} = \mathbf{f}(\mathbf{m}^{(k)}) + \varepsilon^{(k)}$ for "representative" members $\mathbf{m}^{(k)}, \varepsilon^{(k)}$ of the posterior probability density $\sigma(\mathbf{m}|\mathbf{d})$ and of the experimentally determined measurement uncertainty given in (3.2). One would then test that the actual measurements $\mathbf{d}$ are compatible with the statistics of the collection $\mathbf{d}^{(k)}$.

For our experiments, such a validation did not indicate any major issues. If this had been an experiment for which we did not know the "true" value of gravity, we would have likely closed the book and declared victory. On the other hand, besides the poor match of $\mathbb{E}(g|\mathbf{d})$ to the known value $\hat{g}$, there was one additional clue that something might still be missing and that illustrates the importance of experience and intuition in using experimental data. We will discuss this in the following section.

6. Adding a Third Parameter: Start Time. While extracting data frame by frame for the distance the body fell, we realized that it was not trivial to determine the time the body was released. The time Δt between frames is one thirtieth of a second and the distance the body falls within the first frame after release is $\frac{1}{2}g\Delta t^2 \approx 0.5$ cm. The camera was positioned in such a way that it could see at least 5 m of vertical distance. At the resolution of typical inexpensive video cameras, 0.5 cm is significantly less than one pixel. To compound the problem, video footage is typically compressed and blurred, further reducing the effective resolution.

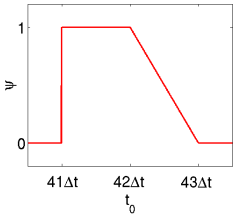
After flipping back and forth between frames many times, the best we could say was that the body was definitely moving between frames 42 and 43 and, possibly, though we could not really determine this with any degree of confidence, between frames 41 and 42. The body was not moving between frames 40 and 41. To put this in mathematical terms, the start time t_0 definitely satisfied $41\Delta t \le t_0 \le 43\Delta t$ and we are inclined to believe that in fact $41\Delta t \le t_0 \le 42\Delta t$.

This is not a strong statement, but it is typical of any physical measurement process: we would need to have the same sort of discussion had we used a higher resolution camera with a higher frame rate. Pressed to choose a start time t_0, we determined that frame 42 was our best guess and, consequently, chose $t_0 = 42\Delta t$ for the computations shown in section 4. Furthermore, by using various values for t_0 in computing the results of section 4, we realized that $\sigma(\mathbf{m}|\mathbf{d})$ varies dramatically with t_0. In other words, the *sensitivity* of $\sigma(\mathbf{m}|\mathbf{d})$ with respect to t_0 is large, indicating that we would need to know t_0 accurately to determine $\sigma(\mathbf{m}|\mathbf{d})$ accurately.

The difficulties we encountered in the choice of start time and the sensitivity of the results with respect to this value make it clear that in reality t_0 is just another parameter in our model (2.3)–(2.4), and one that we know only up to some uncertainty. We should therefore treat it using the techniques for uncertain parameters we introduced in sections 3.2 and 3.3: We let the set of model parameters $\mathbf{m}$ be $\mathbf{m} = \{g, C, t_0\}$, define a prior probability density $\rho_M(\mathbf{m})$ as in (3.1), define the likelihood $\rho_D(\mathbf{d}|\mathbf{m})$ as in (3.3), and combine them into the posterior probability density $\sigma(\mathbf{m}|\mathbf{d}) = \sigma(g, C, t_0|\mathbf{d}) = k\ \rho_M(\mathbf{m})\ \rho_D(\mathbf{d}|\mathbf{m})$ for all three parameters as in (3.4). In the process of evaluating ρ_D we have to compute $\mathbf{f}(\mathbf{m})$, which is now a function that predicts fall distances at times t_i for the values of g, C, and t_0 given by the tuple $\mathbf{m}$. $\mathbf{f}(\mathbf{m})$ can be computed in exactly the same way as before.

The only step that differs significantly is the definition of the prior probability $\rho_M(\mathbf{m})$. We assume that t_0 is independent of g and C and has a pdf given by the function

$$\psi(t_0) = \begin{cases} 1 & \text{for } 41\Delta t \le t_0 \le 42\Delta t, \\ 1 - \frac{t_0 - 42\Delta t}{\Delta t} & \text{for } 42\Delta t \le t_0 \le 43\Delta t, \\ 0 & \text{for } t_0 < 42\Delta t \text{ or } 43\Delta t < t_0 \end{cases}$$



that encodes our best knowledge of the start time. The prior pdf for $\mathbf{m}$ is then

$$\rho_M(\mathbf{m}) = \rho_M(g, C, t_0) = \kappa_M \ \chi_{\left[0,20\frac{\mathrm{m}}{\mathrm{s}^2}\right]}(g) \ \ \chi_{[0,0.5\mathrm{m}^{-1}]}(C) \ \ \psi(t_0), \tag{6.1}$$

where κ_M is another normalization constant different from the one chosen in (3.1). Certainly, other forms for $\psi(t_0)$ could be chosen, leading to slightly different results, but any choice would have the same general structure to be compatible with the discussion at the beginning of this section.

With all these pieces in place, we can evaluate the new posterior density $\sigma(\mathbf{m}|\mathbf{d})$ indicating the relative probability of different values of g, C, and t_0. From this, we compute expected values through (1.2), using quadrature on a mesh in $\mathbf{m}$ space to replace integrals, and obtain

$$\begin{aligned} \mathbb{E}(g|\mathbf{d}) &= 8.64\frac{\mathrm{m}}{\mathrm{s}^2}, \\ \mathbb{E}(C|\mathbf{d}) &= 0.106\mathrm{m}^{-1}, \\ \mathbb{E}(t_0|\mathbf{d}) &= 42\Delta t - 0.005\mathrm{s}. \end{aligned} \tag{6.2}$$

Here, we have used 601, 141, and 101 mesh points for quadrature for the ranges $6\mathrm{m/s}^2 \le g \le 12\mathrm{m/s}^2$, $-0.2\mathrm{m}^{-1} \le C \le 0.5\mathrm{m}^{-1}$, and $42\Delta t - 0.05\mathrm{s} \le t_0 \le 42\Delta t + 0.05\mathrm{s}$, respectively. Again, we have verified that this mesh is sufficiently fine that quadrature errors are negligible.

We see that our results indicate that the most likely starting time is one-sixth of a frame before frame 42. However, we also find that $\mathrm{stddev}(t_0|\mathbf{d}) = 0.015\mathrm{s}$; i.e., we know the start time only with a significant uncertainty of almost one half of a frame. To put this standard deviation in context, consider that the prior probability by itself has expectation values and standard deviations equal to

$$\begin{aligned} \mathbb{E}(t_0) &= \left[\int_{-\infty}^{\infty} \psi(t_0)\, \mathrm{d}t_0\right]^{-1} \int_{-\infty}^{\infty} t_0\, \psi(t_0)\, \mathrm{d}t_0 = \left(42 - \frac{2}{9}\right)\Delta t \\ &\approx 42\Delta t - 0.0074\mathrm{s}, \end{aligned}$$

$$\begin{aligned} \mathrm{stddev}(t_0) &= \sqrt{\left[\int_{-\infty}^{\infty} \psi(t_0)\, \mathrm{d}t_0\right]^{-1} \int_{-\infty}^{\infty} (t_0 - \mathbb{E}(t_0))^2\, \psi(t_0)\, \mathrm{d}t_0} = \sqrt{\frac{37}{162}}\Delta t \\ &\approx 0.48\Delta t \approx 0.016\mathrm{s}. \end{aligned}$$

In other words, our measurement data are only moderately successful in telling us anything more about the start time than what we already knew and encoded in our prior probability density.

On the other hand, the addition of t_0 to the set of parameters to be determined has a significant impact on the parameters we recover. In particular, the posterior mean of g is farther from the true value $\hat{g}$ from (4.2). This does not sound like an improvement at first, but we can show that allowing t_0 to be uncertain leads to a much *larger* uncertainty in g compared to the results of section 4. To see this, we can again compute the marginal posterior density for gravity:

$$\sigma(g|\mathbf{d}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma(g, C, t_0|\mathbf{d})\, \mathrm{d}t_0\, \mathrm{d}C.$$

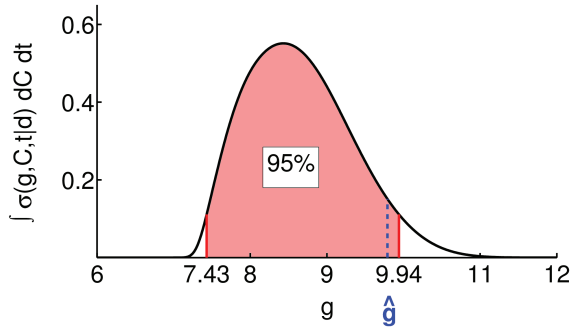


Fig. 6.1 *Marginal posterior density for gravity $\sigma(g|\mathbf{d})$. Indicated in red is the smallest interval of 95% credibility; the true value of gravity $\hat{g} = 9.7935$ is shown in blue.*

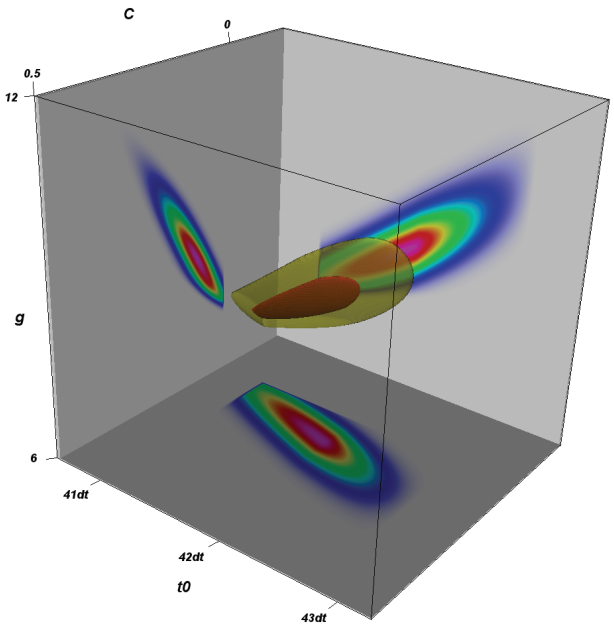


Fig. 6.2 *Isocontours for the 50 (red) and 95% (yellow) credibility levels and marginal probabilities (projected on the sides of the box) of $\sigma(g, C, t_0|\mathbf{d})$. Since the prior probability $\rho_M(\mathbf{m})$ assigns zero probability to values $C < 0$ and $t_0 < 41\Delta t$, the credible regions have sharp edges at these threshold values.*

A plot of $\sigma(g|\mathbf{d})$ is shown in Figure 6.1. This time, $\hat{g}$ lies well within the now much wider 95% credibility interval $[7.43\text{m/s}^2, 9.94\text{m/s}^2]$. As a consequence of introducing uncertainty in t_0, the true gravity constant $\hat{g}$ is now a quite plausible value for g.

To gain a further understanding of the structure of the posterior probability density, we plot $\sigma(g, C, t_0|\mathbf{d})$ in the same way as we did in Figure 4.1 for the case of two parameters. Figure 6.2 shows similar results for all three parameters. Visualizing

three- or higher-dimensional probability densities is notoriously difficult. One strategy, used in the picture, is to show the boundaries of the *credible regions* $I_{50\%}$ and $I_{95\%}$. A credible region I_r is defined by a threshold value s_r as that volume in $\mathbb{R}^3$ so that

$$\sigma(\mathbf{m}|\mathbf{d})|_{I_r} \geq s_r, \qquad \sigma(\mathbf{m}|\mathbf{d})|_{\mathbb{R}^3 \setminus I_r} < s_r, \qquad \int_{I_r} \sigma(\mathbf{m}|\mathbf{d}) \, \mathrm{d}^3\mathbf{m} = r.$$

In other words, the credible region of level r is the region of highest probability density, so that the combined probability for all points lying inside it is r. $I_{50\%}$ can be interpreted as follows: due to measurement uncertainties, we do not know where exactly the "true" parameter $\mathbf{m}$ lies, but we know with 50% probability that it is within the region $I_{50\%}$.

To further illustrate the shape and three-dimensional structure of the probability density σ, the sides of the box shown in Figure 6.2 depict the *marginal* probabilities obtained by integrating over the direction perpendicular to a plane; for example, the back left side of the box shows

$$\sigma(g, C|\mathbf{d}) = \int_{-\infty}^{\infty} \sigma(g, C, t_0|\mathbf{d}) \, \mathrm{d}t_0.$$

This is the probability density we are actually interested in and for which we initially performed the physical experiment, as we only care about g and C.

7. Conclusions and Outlook. Traditionally, parameter estimation was concerned with finding single estimates of the parameters of systems—for example, the often-used least squares fit of parameters to data. This gives useful information about parameters, but it is frequently difficult to assess how certain we can be of these estimates beyond, say, the computation of standard deviations based on linearized models and Gaussian noise assumptions. Bayesian inversion as discussed here provides an alternative viewpoint: Our discussions above illustrate how we can determine probability densities of parameters from measurements obtained in physical experiments, carefully accounting for our prior knowledge as well as our best guess of uncertainties in our measurements. The posterior probability density resulting from Bayesian inversion provides significantly more information than any single estimator of the parameters.

This additional information is important. For example, assume that we have obtained an estimate for the density and viscosity of air from an experiment. If our goal is to verify that a particular wing design provides enough lift for a plane to fly, then we could do a numerical simulation with these particular values of the parameters and confirm that it indeed does. On the other hand, what if the experimental setup was able to give us only a rough estimate of the parameters, and the values we used in the numerical simulation were in fact inaccurate? In that case, the conclusions from the simulation would be worthless. This is avoidable when inverting parameters using the Bayesian framework discussed in this paper: There, the result of inversion is not a single set of parameters but rather a probability distribution on the space of parameters. Consequently, if our goal is to predict outcomes, predictions will be uncertain when there is uncertainty in the parameters used as input, even if the model is exact. As a result, the best that can be done is to provide a probability distribution for predicted *output* quantities, and, in the example, a probability that the plane will in fact fly. Producing such probability distributions in the case of many input parameters and output quantities is a difficult task, and it is the essence of the mathematical field of *uncertainty quantification* [3, 18, 22, 28].

We should also step back and ask whether our Bayesian inversion of measurement data was the best we could have done to determine a probability distribution for the parameters in our model. We want to constrain the parameters as much as possible; i.e., we want to measure in such a way that the posterior probability density is as localized as possible. Ideally, the probability distributions $\sigma(g, C|\mathbf{d})$ or $\sigma(g, C, t_0|\mathbf{d})$ should be Dirac delta functions, but, given the measurement data we have, we have to accept the probability density produced by the Bayesian inversion formalism as it is. To get a more localized density we have to change what we measure. The field of *optimal experimental design* [1, 6, 9, 19, 20] deals with the question of how to measure things so as to obtain the best estimates of the parameters (as defined by a particular criterion) given the available resources and experimental constraints.

To illustrate the importance of experimental design, let us consider what would happen if we had only been able to determine the distance the body had fallen in the first five or last five frames shown in Figure 3.3, respectively. Five data points should be enough to determine two parameters, but the uncertainties are generally larger if we have less data available. Indeed, this is what we see in the probability densities shown in Figure 7.1: we show the results of following the same procedure as in section 4 to find $\sigma(g, C|\mathbf{d})$, but using only the first (left panel) or last five (right panel) data points. These results are easily interpreted. If we use only the first five data points, then our posterior pdf is almost uninformative: it is zero for negative values of the friction coefficient (as a result of our choice of prior probability for this parameter), but apart from that, almost constant and very broad in the direction of the gravity constant g. In effect, the posterior density σ has significant values in the entire range shown. This makes sense because at early times, the body is still too slow for friction to play a significant role, and what we measure is almost completely independent of the air friction coefficient. Second, the error bars are so large compared to the measured values at early times that wildly different values of g lead to predicted falling distances $z(t_i)$ that are all still compatible with the error bars.

On the other hand, if we only have measurements from the last five video frames, then gravity and friction are already close to a balance and the body's speed is approximately equal to $v_\infty = \sqrt{g/C}$. As a result, we are no longer able to determine g and C independently; we can only determine their ratio g/C, which is constant along lines in g-C-space. This explains the elongated shape—i.e., the high degree of correlation—of the recovered probability density.

The lesson from this is that the choice of *what, where, and how* we measure has a large impact on how accurately we can determine the parameters. This realization allows us to reveal our reasons for choosing the setup of the experiment from which we obtained the data used in this article. We performed many experiments, dropping a variety of objects from different heights. Most of the objects we tried were either too heavy compared to their size (balls, beanbags) or too light (the paper bag by itself, air balloons). In the first case, our experiment was too short to see air friction act (see the left panel of Figure 7.1) or the object was so light that it almost instantly assumed its equilibrium velocity (see the right panel of Figure 7.1). Other objects, such as the stuffed animal Mr. Tiger by himself (without the brown paper bag), had such odd shapes that it was impossible to determine how far they had fallen in a given frame. Only a judicious choice of experiment from the many we performed allowed us to invert data in such a way that the posterior probability density $\sigma(g, C|\mathbf{d})$ was reasonably localized and therefore allowed the determination of parameters without unduly large uncertainties, underlining the point made at the beginning of this paragraph.

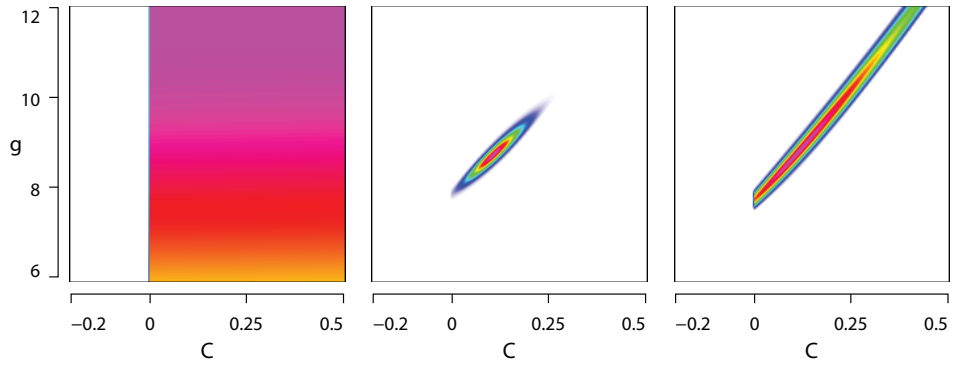


Fig. 7.1 *$\sigma(g, C|\mathbf{d})$ for the case where we use only the first five data points of Figure* 3.3 *(left) or only the last five data points (right). The center plot was obtained using all data points and matches Figure* 4.1.

Acknowledgment. The authors are enormously grateful to Luis Tenorio, who went far beyond the call to make this paper happen. Thank you, Luis!

REFERENCES

[1] A. C. Atkinson and A. N. Donev, *Optimum Experimental Design*, Clarendon Press, Oxford, 1992.
[2] G. Beutler, *Earth gravity field from space—from sensors to earth sciences: Closing remarks*, Space Sci. Rev., 110 (2004), pp. 359–368.
[3] L. T. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, and K. Wilcox, eds., *Large-scale Inverse Problems and Quantification of Uncertainty*, Wiley Series in Computational Statistics, Wiley, Chichester, UK, 2011.
[4] Canon Inc., *Canon ZR*70 *MC/ZR*65 *MC/ZR* 60 *Instruction Manual*, 2003. http://ec1.images-amazon.com/media/i3d/01/A/man-migrate/MANUAL000000857.pdf.
[5] J. B. Carlin and T. A. Louis, *Bayesian Methods for Data Analysis*, 3rd ed., Chapman & Hall, Boca Raton, FL, 2008.
[6] K. Chaloner and I. Verdinelli, *Bayesian experimental design: A review*, Statist. Sci., 10 (1995), pp. 273–304.
[7] S. Chib and E. Greenberg, *Understanding the Metropolis-Hastings algorithm*, Amer. Statist., 49 (1995), pp. 327–335.
[8] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, 1996.
[9] E. Haber, Z. Magnant, C. Lucero, and L. Tenorio, *Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems*, Comput. Optim. Appl., 52 (2012), pp. 293–314.
[10] S.-C. Han, C. K. Shum, and A. Braun, *High-resolution continental water storage recovery from low-low satellite-to-satellite tracking*, J. Geodyn., 39 (2005), pp. 11–28.
[11] W. A. Heiskanen and F. A. Vening Meinesz, *The Earth and Its Gravity Field*, McGraw-Hill, New York, 1958.
[12] H. Jeffreys, *The determination of the earth's gravitational field*, Monthly Notices of the Royal Astronomical Society, Geophysical Supplement, 5 (1941), pp. 1–22.
[13] H. Jeffreys, *The determination of the earth's gravitational field (second paper)*, Monthly Notices of the Royal Astronomical Society, Geophysical Supplement, 5 (1943), pp. 55–66.
[14] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, 2006.
[15] W. M. Kaula, *Statistical and harmonic analysis of gravity*, J. Geophys. Res., 64 (1959), pp. 2401–2421.
[16] A. C. Kermode, *Mechanics of Flight*, Prentice Hall, Pearson, Harlow, UK, 2006.
[17] D. Kincaid and W. Cheney, *Numerical Analysis: Mathematics of Scientific Computing*, 3rd ed., Brooks/Cole, Pacific Grove, CA, 2002.

[18] D. Kurowicka and R. M. Cooke, *Uncertainty Analysis with High Dimensional Dependence Modelling*, Wiley Series in Probability and Statistics, Wiley, Chichester, UK, 2006.
[19] A. Pázman, *Foundations of Optimum Experimental Design*, D. Reidel, Dordrecht, The Netherlands, 1986.
[20] F. Pukelsheim, *Optimal Design of Experiments*, John Wiley & Sons, New York, 1993.
[21] A. H. Shapiro, *Shape and Flow: The Fluid Dynamics of Drag*, Anchor Books, 1961.
[22] P. B. Stark and L. Tenorio, *A primer of frequentist and Bayesian inference in inverse problems*, in Large-scale Inverse Problems and Quantification of Uncertainty, L. T. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, and K. Wilcox, eds., Wiley Series in Computational Statistics, Wiley, Chichester, UK, 2011.
[23] A. Tarantola, *Inverse Problem Theory*, Elsevier, Amsterdam, New York, 1987.
[24] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.
[25] L. Tenorio, F. Andersson, M. de Hoop, and P. Ma, *Data analysis tools for uncertainty quantification of inverse problems*, Inverse Problems, 27 (2011), 045001.
[26] Wikipedia, *Earth's gravity*, http://en.wikipedia.org/wiki/Earth%27s_gravity.
[27] Wikipedia, *Gravimetry*, http://en.wikipedia.org/wiki/Gravimetry.
[28] S. F. Wojtkiewicz, M. S. Eldred, R. V. Field, A. Urbina, and J. R. Red-Horse, *Uncertainty quantification in large computational engineering models*, in Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials (SDM) Conference, 2001, Seattle WA, 2001, AIAA 2001-1455, pp. 1–11.