Springer Series in Reliability Engineering

Dana Kelly
Curtis Smith

Bayesian Inference for Probabilistic Risk Assessment

A Practitioner's Guidebook

Springer

Springer Series in Reliability Engineering

For further volumes:
http://www.springer.com/series/6917

Dana Kelly · Curtis Smith

Bayesian Inference for Probabilistic Risk Assessment

A Practitioner's Guidebook

Springer

Dana Kelly
Idaho National Laboratory (INL)
PO Box 1625
Idaho Falls, ID 83415-3850
USA
e-mail: Dana.Kelly@inl.gov

Curtis Smith
Idaho National Laboratory (INL)
PO Box 1625
Idaho Falls, ID 83415-3850
USA
e-mail: Curtis.Smith@inl.gov

ISSN 1614-7839
ISBN 978-1-84996-186-8 e-ISBN 978-1-84996-187-5
DOI 10.1007/978-1-84996-187-5
Springer London Dordrecht Heidelberg New York

British Library Cataloguing in Publication Data
A Catalogue record for this book is available from the British Library

Cover design: eStudio Calamar, Berlin/Figueres

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book began about 23 years ago, when one of the authors encountered a formula in a PRA procedure for estimating a probability of failure on demand, p. The formula was not the obvious ratio of number of failures to number of demands; instead it looked like this:

$$\tilde{p} = \frac{x + 0.5}{n + 1}$$

Upon consulting more senior colleagues, he was told that this formula was the result of "performing a Bayesian update of a noninformative prior." Due to the author's ignorance of Bayesian inference, this statement was itself quite noninformative.

And so began what has become a career-long interest in all things Bayesian. Both authors have indulged in much self-study over the years, along with a few graduate courses in Bayesian statistics, where they could be found. Along the way, we have developed training courses in Bayesian parameter estimation for the U.S. Nuclear Regulatory Commission and the National Aeronautics and Space Administration, and we continue to teach the descendants of these courses today. We have also developed and presented workshops in Bayesian inference for aging models, and written a number of journal articles and conference papers on the subject of Bayesian inference, all from the perspective of practicing risk analysts.

After having developed a Bayesian inference guidebook for NASA, a guidebook on Bayesian inference for time-dependent problems for the European Commission, and an update of a paper on Bayesian parameter estimation written in the 1990s, we were approached by Springer with the idea of writing a textbook. There is an ever-increasing number of Bayesian inference texts on the market, due in large part to the growth in computing power and the accompanying popularization of the Markov Chain Monte Carlo (MCMC) techniques we employ herein. Many, if not most of these texts are written at the level of an advanced undergraduate or beginning statistics graduate student. In other words, the available references are generally beyond the level of the typical risk analyst in the field,

who is most often an engineer, and who may have had at most a course or two in probability and statistics along the way, typically from a frequentist perspective.

Having struggled without success to find a suitable text for our courses over the years, we wanted to write a text that would be accessible to a majority of practicing risk analysts. We wanted to employ the modern technique of MCMC, which can handle a wide range of what were once intractable problems. Having followed the development of BUGS since its inception in the 1990s, we decided to write the text around this software. There are other choices of software for this type of analysis, many of which are also free and open-source. JAGS (Just Another Gibbs Sampler) is one whose syntax is very similar to that of BUGS. The R software package has a number of MCMC routines available, as well as packages that interface with BUGS and JAGS, allowing them to be run in "batch mode." R also has packages for processing the output of BUGS and JAGS, including convergence diagnostics and graphics. The Python language has PyMC, which is the focus of much advanced development efforts these days. We encourage the interested reader to explore these software packages, along with others that will come along in the future.

BUGS has evolved into OpenBUGS, and most of the examples in this text were solved using OpenBUGS 3.1.2. At the time this text went to press, Ver. 3.2.1 was released. This version has significant enhancements not covered in this text, most notably ReliaBUGS, which includes a variety of specialized distributions used in advanced reliability analysis. It's probably time to begin work on a second edition!

We hope to remove some of the mystery that seems to surround formulas like the one above, and to make plain often-heard incantations such as "Bayesian update of a noninformative prior." For after all, Bayesian inference is, we feel, much more straightforward than its frequentist alternative, especially in the interpretation of its results. There is one formula, Bayes' Theorem, which underlies all that is done, and understanding the component parts of this theorem is the key to just about everything else, including the specialized jargon that has accumulated in Bayesian inference, as in every other field in science and engineering.

We begin the text in Chapters 1 and 2 with the motivation for using Bayesian inference in risk assessment, and provide a general overview before moving into the most commonly encountered risk assessment problems in Chapter 3, those involving a single parameter in an aleatory model.

Chapter 4 introduces the too-often overlooked subject of model checking, and illustrates a number of checks, both qualitative graphical ones and quantitative checks based on the posterior predictive distribution.

Chapter 5 introduces more complicated aleatory models in which there is a monotonic trend in the parameters of the commonly used binomial and Poisson distributions.

Chapter 6 discusses MCMC convergence from a practical point of view. In principle, convergence to the joint posterior distribution can be problematic; however, in risk assessment convergence is rarely an issue except in the population

variability models discussed in Chapter 7. However, if there is more than one parameter involved, the prudent analyst will always check for convergence.

Chapter 8 introduces more complex models for random durations, covering inference for the Weibull, lognormal, and gamma distributions as aleatory models. It also introduces penalized likelihood criteria that can be used to select among candidate models, focusing on the deviance information criterion (DIC) that is calculated by OpenBUGS.

Chapters 8 and 9 together describe time-dependent aleatory models that are encountered when modeling the infant mortality and especially the aging portions of the famous "bathtub curve." Chapter 9 covers inference for the so-called renewal process, where failed components are replaced with new ones, or repairs restore the component to a good-as-new state. Chapter 8 covers the situation where repair is same-as-old; rather than reincarnating a failed component, the failed component is merely resuscitated. Chapter 9 also discusses some useful graphical checks for exploratory data analysis.

Chapter 10 turns to analysis of cases where the observed data are uncertain in some way, perhaps because of censoring, inaccurate record-keeping, or other reasons. The Bayesian framework can handle these kinds of uncertainties in a very straightforward extension of the case without such uncertainty.

Chapter 11 introduces regression models, in which additional information in the form of observable quantities such as temperature can enhance the simpler aleatory models used in earlier chapters.

Chapter 12 describes inference at multiple levels of a system fault tree. For example, we might have information on the overall system performance, but we might also have subsystem and component-level information.

Chapter 13 closes the text with a selection of problems, which are generally of a more specialized or advanced nature. These include an introduction to inference for extreme value processes, such as might be employed to model an external flooding hazard, an introduction to treatment of expert opinion in the Bayesian framework, specification of a prior distribution in OpenBUGS that is not one of the built-in choices, and Bayesian inference for the parameters of a Markov model, the last illustrating the ability to numerically solve systems of ordinary differential equations within OpenBUGS, while simultaneously performing Bayesian inference for the parameter values in these equations.

We apologize in advance for the errors that will inevitably be found, and hope that they will not stand in the way of learning.

Dana Kelly
Curtis Smith

Contents

Chapter 1
Introduction and Motivation

1.1 Introduction

The focus for applications in this book is on those related to probabilistic risk analysis (PRA). As discussed in Siu and Kelly [1], PRA is an analysis of the frequency and consequences of accidents in an engineered system. This type of analysis relies on probabilistic (i.e., predictive) models and associated data. Because of PRA's focus on low-frequency scenarios, often involving the failure of highly reliable equipment, empirical data are often lacking. Bayesian inference techniques are useful in such situations because, unlike frequentist statistical methods, Bayesian techniques are able to incorporate non-empirical information. Furthermore, from a practical perspective, Bayesian techniques, which represent uncertainty with probability distributions, provide a ready framework for the propagation of uncertainties through the risk models, via Monte Carlo sampling.

There are even more advantages in adopting a Bayesian inference framework in PRA. For example, in data collection and evaluation, we strive for the situation where the observed data are known with certainty and completeness. Unfortunately, for a variety of reasons, reality is messier in a number of ways with respect to observed data. For example, one may not always be able to ascertain the exact number of failures of a system or component that have occurred, perhaps because of imprecision in the failure criterion, record keeping, or interpretation. Further, when there are multiple failure modes to track (e.g., fails during standby versus fails during a demand), one may not be certain for which failure mode to count a specific instance of a failure.

Even though an estimate of a component's failure rate or failure probability is required for use in the PRA, a detailed analysis and data gathering effort is not always possible for every part/assembly. Consequently, PRA must use the available data and information as efficiently as possible while providing representative uncertainty characterizations—it is this drive to provide probability distributions

D. Kelly and C. Smith, *Bayesian Inference for Probabilistic Risk Assessment*,
Springer Series in Reliability Engineering, DOI: 10.1007/978-1-84996-187-5_1,

representing what is known about elements in the risk analysis that leads us to Bayesian inference.

The hardware modeling element of the PRA has typically relied on operational information and data, coupled with Bayesian inference techniques, to quantify performance. This analysis of performance uses the Bayesian approach in order to escape some of the problems associated with frequentist estimates, including:

- If data are sparse, frequentist estimators, such as "maximum likelihood estimates," can be unrealistic (e.g., zero);
- Propagating frequentist interval estimates, such as confidence intervals, through the PRA model is difficult;
- Frequentist methods are sometimes of an *ad hoc* nature;
- Frequentist methods cannot incorporate "non-data" information into the quantification process, other than in an *ad hoc* manner.

The last limitation of frequentist inference listed above highlights a key attribute of Bayesian methods, namely the ability to incorporate qualitative information (i.e., evidence) into the parameter estimates. Unlike frequentist inference, which focuses solely on "data," the Bayesian approach to inference can bring to bear all of what is known about a process, *including* empirical data.

1.2 Background for Bayesian Inference

The Bayesian (or Bayes–Laplace) method of probabilistic induction has existed since the late 1700s [2, 3]. Laplace, starting in 1772, performed the first quantitative Bayesian inference calculations. The application then was inferring the mass of planets such as Jupiter and Saturn using astronomical observations, along with simple (i.e., uniform) prior distributions, but a rather complicated stochastic model [4]. Unfortunately, the Bayesian mathematics for Laplace's problem was quite complicated, primarily due to his selection of a particular stochastic model (a double exponential, also known today as a Laplace distribution).

Later, in 1809, Gauss popularized the normal (or Gaussian) distribution. Laplace, having been made aware of this new stochastic model, returned in 1812 to his previously intractable problem of inference about the mass of Saturn. To perform his Bayesian calculation, he used:

- Data (orbital information on the satellite Callisto).
- A prior (the uniform distribution).
- A less complicated stochastic model (the normal distribution).

What Laplace calculated was a posterior probability distribution for the mass of Saturn. He published his results in the *Théorie Analytique des Probabilités*, representing the first successful quantitative application of Bayes' Theorem.

Today, the Bayesian approach to inference is employed in a very wide variety of domains for many different stochastic modeling situations. For example, two

widely used stochastic models (both in and outside of PRA) are the Poisson and binomial, representing different processes:

- Examples of Poisson processes.
  - Counting particles, such as neutrons, in a second.
  - Number of (lit) lights failing over a month.
  - Arrival of customers into a store on a Monday.
  - Large earthquakes in a region over a year.
  - HTTP requests to a server during a day.
- Examples of Bernoulli processes.
  - Tossing a coin to see if it comes up heads.
  - Starting a car to see if it will start.
  - Turning on a light to see if will turn on.
  - Launching a rocket to see if it will reach orbit.

The basis of many traditional PRAs is event tree and fault tree models (deterministic models), which logically relate the occurrence of low-level events to a higher-level event (e.g., an initiating event followed by multiple safety system failure events may lead to an undesired outcome). The occurrence of initiating events and system failures (or just "events") in the fault trees and event trees are modeled probabilistically, and the associated probabilistic models each contain one or more parameters, whose values are known only with uncertainty. The application of Bayesian methods to estimate these parameters, with associated uncertainty, uses all available information, leading to informed decisions based upon the applicable information at hand.

Most PRAs require different types of "failure models" to quantify the risk portion of the analysis. These models include the following:

- Failure of a component to change state on demand.
- Failure in time of an operating component.
- Rate of aging for a passive component.
- Failure (in standby) of an active component while in a quiescent period.
- Downtime or unavailability due to testing.
- Restoration of a component following a failure.

We describe Bayesian inference for these and other models in this book, using modern computational tools.

1.3 An Overview of the Bayesian Inference Process

In Chap. 2, we will introduce Bayes Theorem, which according to the theory of subjective probability, is the only way in which an analyst whose probabilities obey the axioms of probability theory can update his or her state of knowledge [5]. The general procedure for performing Bayesian inference is:

1. Specify an *aleatory model*[1] for the process being represented in the PRA (e.g., failure of component to change state on demand).
2. Specify a *prior distribution* for parameter(s) in this model, quantifying epistemic uncertainty, that is, quantifying a state of knowledge about the possible parameter values.
3. Observe *data* from or related to the process being represented.
4. Update the prior to obtain the *posterior distribution* for the parameter(s) of interest.
5. Check validity of the aleatory model, data, and prior.

We follow this process to make inferences, that is, to estimate the probability that a model, parameter, or hypothesis is reasonable, conditional on all available evidence. As part of describing the process of Bayesian inference, we used several terms such as "data," "aleatory," and "epistemic." In this text, we attach specific meanings to key terms for which confusion often exists:

Data Distinct observed values of a physical process. Data may be factual or not. For example they may be subject to uncertainties, such as imprecision in measurement, truncation, and interpretation errors. Examples of data include the **number** of failures during part testing, the **times** at which a tornado has occurred within a particular area, and the **time** it takes to repair a failed component. In these examples, the observed item is bolded to emphasize that data are observable. An aleatory model is used in PRA to model the process that gives rise to data

Information The result of evaluating, processing, or organizing data and information in a way that adds to knowledge. Note that information is not necessarily observable; only the *subset* of information referred to as data is observable. Examples of information include a calculated estimate of failure probability, an expert's estimate of the frequency of tornados occurring within a particular area, and the distribution of a repair rate used in an aleatory model for the restoration time of a failed component

Knowledge What is known from gathered information

Inference The process of obtaining a conclusion based on what one knows.

To evaluate data in an inference process, we must have a "model of the world" (or simply "model") that allows us to translate observable events into information [6, 7]. Within this framework, there are two fundamental types of model

[1] Also referred to synonymously as a "stochastic model," "probabilistic model," or "likelihood function."

abstractions, aleatory and deterministic. The term "aleatory" refers to the stochastic nature of the outcome of a process. For example, flipping a coin, testing a part, predicting tornadoes, rolling a die, etc., are typically (chosen to be) modeled as aleatory processes. In the case of flipping a coin, the observable stochastic data are the outcomes of the coin flips (heads or tails).

Since probabilities are not observable quantities, we do not have a model of the world directly for probabilities. Instead, we rely on aleatory models (e.g., a Bernoulli[2] model in the case of tossing a coin) to infer probabilities for observable outcomes (e.g., two heads out of three tosses of the coin).

Aleatory Pertaining to stochastic (non-deterministic) events, the outcome of which is described using probability. From the Latin *alea* (a game of chance or a die)

Deterministic Pertaining to exactly predictable (or precise) processes, the outcome of which is known with certainty if the inputs are known with certainty. As the antithesis of aleatory, this is the type of model most familiar to scientists and engineers and includes relationships such as $V = IR$, $E = mc^2$, $F = ma$, $F = G\, m_1\, m_2/r^2$

In PRA, we employ both aleatory and deterministic models. In these models, even ones that are deterministic physical models, such as thermal–hydraulic models used to derive system success criteria, many of the parameters are themselves imprecisely known, and therefore are treated as uncertain variables. To describe this second type of uncertainty (with aleatory uncertainty being the first kind), PRA employs the concept of epistemic uncertainty.

Epistemic Pertaining to the degree of knowledge about models and their parameters. From the Greek *episteme* (knowledge)

Whether we use an aleatory model (e.g., Bernoulli process) or a deterministic model (e.g., $F = ma$), if any parameter in the model is imprecisely known, then there is epistemic uncertainty associated with the output of that model. It is the goal of this book to demonstrate how to combine data and information with applicable models and, via Bayesian inference, enhance our knowledge for PRA applications.

[2] A Bernoulli *trial* is an experiment whose outcomes can be assigned to one of two possible states (e.g., success/failure, heads/tails, yes/no), and mapped to two values, such as 0 and 1. A Bernoulli *process* is obtained by repeating the same Bernoulli trial, where each trial is independent of the others. If the outcome given for the value "1" has probability p, it can be shown that the summation of n Bernoulli trials is binomially distributed $\sim$ binomial (p, n).

References

1. Siu NO, Kelly DL (1998) Bayesian parameter estimation in probabilistic risk assessment. Reliab Eng Syst Saf 62:89–116
2. Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Philos Trans R Soc 53:370–418
3. Laplace PS (1814) A philosophical essay on probabilities. Dover Publications, New York (reprint 1996)
4. Stigler SM (1986) The history of statistics: the measurement of uncertainty before 1900. Belknap Press, Cambridge
5. de Finetti B (1974) Theory of probability. Wiley, New York
6. Winkler RL (1972) An introduction to Bayesian inference and decision. Holt, Rinehart, and Winston, New York
7. Apostolakis GE (1994) A commentary on model uncertainty. In: Mosleh A, Siu N, Smidts C, Lui C (eds) Proceedings of workshop I in advanced topics in risks and reliability analysis, model uncertainty: its characterization and quantification. NUREG/CP-0138, U.S. Nuclear Regulatory Commission, Washington

Chapter 2
Introduction to Bayesian Inference

2.1 Introduction

As discussed in Chap. 1, Bayesian statistical inference relies upon Bayes' Theorem to make coherent inferences about the plausibility of a hypothesis.

Observable data is included in the inference process. In addition, other information about the hypothesis is included in the inference. Consequently, in the Bayesian inference approach, probability quantifies a state of knowledge and represents the plausibility of an event, where "plausibility" implies apparent validity. In other words, Bayesian inference uses probability distributions to encode information , where the encoding metric is a probability (on an absolute scale from 0 to 1).

Note that the use of the word "hypothesis" here should not be confused with classical Neyman-Pearson hypothesis testing. Instead, the types of hypotheses that might be evaluated when performing PRA include:

- The ability of a human to carry out an action when following a written procedure.
- The chance for multiple redundant components to fail simultaneously.
- The frequency of damaging earthquakes to occur at a particular location.
- The chance that a component fails to start properly when demanded.

2.2 Bayes' Theorem

Bayes' Theorem provides the mathematical means of combining information and data, in the context of a probabilistic model, in order to update a prior state of knowledge. This theorem modifies a prior probability, yielding a posterior probability, via the expression:

D. Kelly and C. Smith, *Bayesian Inference for Probabilistic Risk Assessment*,
Springer Series in Reliability Engineering, DOI: 10.1007/978-1-84996-187-5_2,

Table 2.1 Components of *Bayes*' Theorem

Term	Description
$P(H\|D)$	Posterior distribution, which is conditional upon the data D that is known related to the hypothesis H
$P(H)$	Prior distribution, from knowledge of the hypothesis H that is independent of data D
$P(D\|H)$	Likelihood, or aleatory model, representing the process or mechanism that provides data D
$P(D)$	Marginal distribution, which serves as a normalization constant

$$P(H|D) = P(H)\frac{P(D|H)}{P(D)}. \tag{2.1}$$

If we dissect Eq. 2.1, we will see there are four parts (Table 2.1) :

In the context of PRA, where we use probability distributions to represent a state of knowledge regarding parameter values in the PRA models, Bayes' Theorem gives the posterior distribution for the parameter (or multiple parameters) of interest, in terms of the prior distribution, failure model, and the observed data, which in the general continuous form is written as:

$$\pi_1(\theta|x) = \frac{f(x|\theta)\pi_0(\theta)}{\int f(x|\theta)\pi_0(\theta)d\theta} = \frac{f(x|\theta)\pi_0(\theta)}{f(x)}. \tag{2.2}$$

In this equation, $\pi_1(\theta|x)$ is the posterior distribution for the parameter of interest, denoted as θ (note that θ can be a vector). The posterior distribution is the basis for all inferential statements about θ, and will also form the basis for model validation approaches to be discussed later. The observed data enters via the aleatory model, $f(x|\ \theta)$, and $\pi_0(\theta)$ is the prior distribution of θ.

Note that the denominator in Eq. 2.2 has a range of integration that is over all possible values of θ, and that it is a weighted average distribution, with the prior distribution $\pi_0(\theta)$ acting as the weighting function.

In cases where X is a discrete random variable (e.g., number of events in some period of time), $f(x)$ is the probability of seeing exactly x events, unconditional upon a value of θ. If X is a continuous outcome, such as time to suppress a fire, $f(x)$ is a density function, giving the unconditional probability of observing values of X in an infinitesimal interval about x. In a later context, associated with model validation, $f(x)$ will be referred to as the *predictive distribution* for X.

The likelihood function $f(x|\theta)$, or just likelihood, is also known by another name in PRA applications—it is the aleatory model describing an observed physical process. For example, a component failure to operate may be modeled inside a system fault tree by a Poisson process. Or, we may use an exponential distribution to represent fire suppression times. In these cases, there is an inherent modeling tie from the PRA to the data collection and evaluation process—specific aleatory models imply specific types of data. In traditional PRA applications, the aleatory model is most often binomial, Poisson, or exponential, giving rise to data in the form of failures over a specified number of demands, failures over

a specified period of time, and failure times, respectively. Bayesian inference for each of these cases will be discussed in detail in Chap. 3.

The prior distribution, $\pi_0(\theta)$, represents what is known about a parameter θ independent of data generated by the aleatory model that will be collected and evaluated. Prior distributions can be classified broadly as either *informative* or *noninformative*. Informative priors, as the name suggests, contain substantive information about the possible values of θ. Noninformative priors, on the other hand, are intended to let the data dominate the posterior distribution; thus, they contain little substantive information about the parameter of interest.

2.3 A Simple Application of Bayes' Theorem

2.3.1 The Discrete Case

If we toss a coin, can we tell if it is an unfair coin? Specifically, what can the Bayesian approach to inference do to assist in answering this question? The issue that we are concerned with is the possibility of an unfair coin (e.g., a two-headed coin; for now, we will ignore the possibility of a two-tailed coin or a biased coin-tosser to simplify the presentation) being used. Let us jump directly into the Bayes analysis to see how straightforward this type of analysis can be in practice.

First, we note that Bayesian methods rely on three items:

- An aleatory model.
- A prior distribution for the parameter(s) of the aleatory model.
- Data associated with the aleatory model.

As discussed earlier, the prior distribution encodes the analyst's state of knowledge about a hypothesis. In this example, we have two hypotheses (**H**) we are going to consider:

$\mathbf{H_1}$ = we have a fair coin
$\mathbf{H_2}$ = we have an unfair coin

Recall in this example, that an unfair coin implies a two-headed coin. Thus, the probability of heads associated with $\mathbf{H_2}$ would be 1.0 (since we cannot obtain a tail if in fact we have two heads). At this point, we are ready to specify the prior distribution.

Step 1: The Aleatory Model The likelihood function (or aleatory model) representing "random" outcomes (head/tail) for tossing a coin will be assumed to be given by a Bernoulli model:

$$P(D|H_i) = p_i$$

where p_i is the probability of obtaining a head on a single toss conditional upon the ith hypothesis.

Table 2.2 Bayesian inference of one toss of a coin in an experiment to test the hypothesis of a fair coin

Hypothesis	Prior probability	Likelihood	Prior × likelihood	Posterior probability
H_1: fair coin (i.e., the probability of a heads is 0.5)	0.75	0.5	0.375	0.60
H_2: two-headed coin (i.e., the probability of a heads is 1.0)	0.25	1.0	0.250	0.40
	Sum: 1.00		Sum: 0.625	Sum: 1.00

Step 2: The Prior Distribution Knowledge of the "experiment" might lead us to believe there is a significant chance that an unfair coin will be used in the toss. Thus, for the sake of example, let us assume that we assign the following prior probabilities to the two hypotheses:

$$\mathrm{P}(\mathrm{H}_1) = 0.75 \qquad \mathrm{P}(\mathrm{H}_2) = 0.25$$

This prior distribution implies that we think there is a 25% chance that an unfair coin will be used for the next toss. Expressed another way, this prior belief corresponds to odds of 3:1 that the coin is fair.

Step 3: The Data The coin is tossed once, and it comes up heads.

Step 4: Bayesian Calculation to Estimate Probability of a Fair Coin, $P(H_1)$ The normalization constant in Bayes' Theorem, $P(D)$, is found by summing the product of the prior distribution and the aleatory model over all possible hypotheses, which in this example gives

$$P(D) = P(H_1)p_1 + P(H_2)p_2 = (0.75)0.5 + (0.25)1.0 = 0.625$$

where for hypothesis $\mathbf{H_1}$, $p_1 = 0.5$ while for $\mathbf{H_2}$, $p_2 = 1.0$. At this point, we have the aleatory model (as a function of our one data point), the prior distribution, and the normalization constant in Bayes' Theorem. Thus, we can compute the posterior probabilities for our two hypotheses. When we do that calculation, we find:

$$\mathrm{P}(\mathbf{H_1}|\ \text{one toss, data are ``heads''}) = 0.6$$
$$\mathrm{P}(\mathbf{H_2}|\ \text{onetoss, data are ``heads''}) = 0.4$$

The results after one toss are presented in Table 2.2 and show that the posterior probability is the normalized product of the prior probability and the likelihood (e.g., H_1 posterior is 0.375 / 0.625 = 0.60).

What has happened in this case is that the probability of the second hypothesis (two-headed coin) being true has increased by almost a factor of two simply by tossing the coin once and observing heads as the outcome.

As additional data are collected, we can evaluate the impact of the data on our state of knowledge by applying Bayes' Theorem sequentially as the data are

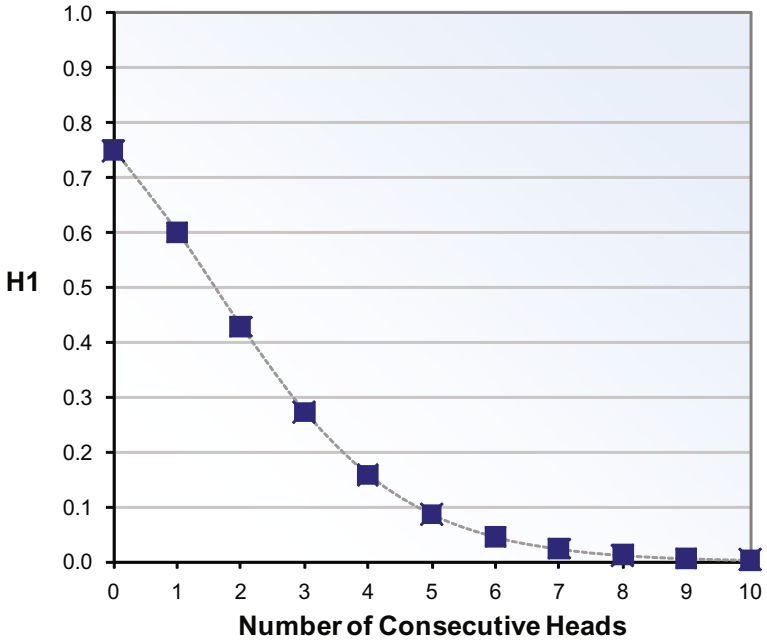

Fig. 2.1 Plot of the posterior probability of a fair coin as a function of the number of consecutive heads observed in independent tosses of the coin

collected. For example, let us assume that we toss the coin j times and want to make inference on the hypotheses (if a head comes up) each time. Thus, we toss ($x = 1, 2, \ldots, j$) the coin again and again independently, and each time the estimate of the probability that the coin is fair changes. We see this probability plotted in Fig. 2.1, where initially (before any tosses) the prior probability of a fair coin (H_1) was 0.75. However, after five tosses where a head appears each time, the probability that we have a fair coin is small, less than ten percent.

2.3.2 The Continuous Case

Let us revisit the example described in Sect. 2.3.1 but employ a continuous prior distribution. The posterior distribution for this example will then be given by:

$$\pi_1(p \,|\, x,\, n) \;=\; \frac{f(x \,|\, p,\, n)\, \pi_0(\theta)}{\int_0^1 f(x \,|\, p,\, n)\, \pi_0(\theta)\; \mathrm{d}\theta}$$

Since we used a Bernoulli aleatory model for the outcome of a coin toss, this leads to a binomial distribution as the aleatory model for the number of heads in n independent coin tosses in which the probability of heads on any toss is p:

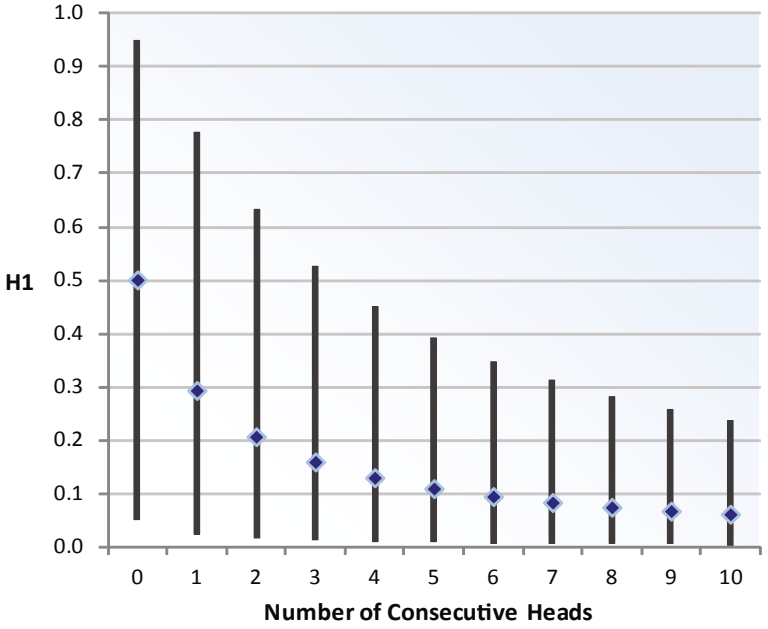

Fig. 2.2 Plot of the posterior probability of a fair coin as a function of the number of consecutive heads observed in independent tosses

$$f(x|p,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

To represent a prior state of ignorance, the prior could be specified as a uniform distribution between $p = 0$ and $p = 1$, or:

$$\pi_0(p) = 1 \qquad (0 \leq p \leq 1)$$

This is an example from the general category of noninformative priors mentioned earlier. We can now rewrite Bayes' Theorem for this example as:

$$\pi_1(p|x,n) = \frac{\theta^x (1-\theta)^{n-x}}{\int_0^1 \theta^x (1-\theta)^{n-x} d\theta} = \frac{\theta^x (1-\theta)^{n-x}}{\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}}$$

It can be shown that the posterior distribution is a beta distribution with parameters:[1]

[1] The uniform distribution in this example is a particular instance of a conjugate prior, to be discussed in more detail in Chap. 3.