

Chapter 11

Introduction to Bayesian Analysis of Hydrologic Variables

Wilson Fernandes and Artur Tiago Silva

11.1 Historical Background and Basic Concepts

According to Brooks (2003) and McGrayne (2011), Bayesian methods date back to 1763, when Welsh amateur mathematician Richard Price (1723–1791) presented the theorem developed by the English philosopher, statistician, and Presbyterian minister Thomas Bayes (1702–1761) at a session of the Royal Society in London. The underlying mathematical concepts of Bayes' theorem were further developed during the nineteenth century as they stirred the interest of renowned mathematicians such as Pierre-Simon Laplace (1749–1827) and Carl Friedrich Gauss (1777–1855), and of important statisticians such as Karl Pearson (1857–1936). By the early twentieth century, use of Bayesian methods declined due, in part, to the opposition of prestigious statisticians Ronald Fisher (1890–1962) and Jerzy Neyman (1894–1981), who had philosophical objections to the degree of subjectivity that they attributed to the Bayesian approach. Nevertheless, prominent statisticians such as Harold Jeffreys (1891–1989), Leonard Savage (1917–1971), Dennis Lindley (1923–2013) and Bruno de Finetti (1906–1985), and others, continued to advocate in favor of Bayesian methods by developing them and prescribing them as a valid alternative to overcome the shortcomings of the frequentist approach.

In the late 1980s, there was a resurgence of Bayesian methods in the research landscape of statistics, due mainly to the fast computational developments of that decade and the increasing need of describing complex phenomena, for which

W. Fernandes (✉)

Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
e-mail: wilson@ehr.ufmg.br

A.T. Silva

CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
e-mail: artur.tiago.silva@tecnico.ulisboa.pt

conventional frequentist methods did not offer satisfactory solutions (Brooks 2003). With increased scientific acceptance of the Bayesian approach, further computational developments and more flexible means of inferences followed. Nowadays it is undisputable that the Bayesian approach is a powerful logical framework for the statistical analysis of random variables, with potential usefulness for solving problems of great complexity in various fields of knowledge, including hydrology and water resources engineering.

The main difference between the Bayesian and frequentist paradigms relates to how they view the parameters of a given probabilistic model. Frequentists believe that parameters have fixed, albeit unknown, true values, which can be estimated by maximizing the likelihood function, for example, as described in Sect. 6.4 of Chap. 6. Bayesian statisticians, on the other hand, believe that parameters have their own probability distribution, which summarize their knowledge, or ignorance, about those parameters. It should be noted that Bayesians also defend that there is a real value (a point) for a parameter, but since it is not possible to determine that value with certainty, they prefer to use a probability distribution to reflect the lack of knowledge about the true value of the parameter. Therefore, as the knowledge of the parameter increases, the variance of the distribution of the parameter decreases. Ultimately, at least in theory, the total knowledge about that parameter would result in a distribution supported on a single point, with a probability equal to one.

Therefore, from the Bayesian perspective, a random quantity can be an unknown quantity that can vary and take different values (e.g., a random variable), or it can simply be a fixed quantity about which there is little or no available information (e.g., a parameter). Uncertainty about those random quantities are described by probability distributions, which reflect the subjective knowledge acquired by the expert when evaluating the probabilities of occurrence of certain events related to the problem at hand.

In addition to the information provided by observed data, which is also considered by the classical school of statistics, Bayesian analysis considers other sources of information to solve inference problems. Formally, suppose that θ is the parameter of interest, and can take values within the parameter space Θ . Let Ω be the available prior information about that parameter. Based on Ω , the uncertainty of θ can be summarized by a probability distribution with PDF $\pi(\theta|\Omega)$, which is called the prior density function or the prior distribution, and describes the state of knowledge about the random quantity, prior to looking at the observed data. If Θ is a finite set, then, it is an inference, in itself, as it represents the possible values taken by θ . At this point, it is worth noting that the prior distribution does not describe the random variability of the parameter but rather the degree of knowledge about its true value.

In general, Ω does not contain all the relevant information about the parameter. In fact, the prior distribution is not a complete inference about θ , unless, of course, the analyst has full knowledge about the parameter, which does not occur in most real situations. If the information contained in Ω is not sufficient, further information about the parameter should be collected. Suppose that the random variable X ,

which is related to θ , can be observed or sampled; prior to sampling X and assuming that the current value of θ is known, the uncertainty about the amount X can be summarized by the likelihood function $f(x|\theta, \Omega)$. It should be noted that the likelihood function provides the probability of a particular sample value x of X occurring, assuming that θ is the true value of the parameter. After performing the experiment, the prior knowledge about θ should be updated using the new information x . The usual mathematical tool for performing this update is Bayes' theorem. Looking back at Eq. (3.8) and taking as reference the definition of a probability distribution, the *posterior* PDF, which summarizes the updated knowledge about θ is given by

$$\pi(\theta|x, \Omega) = \frac{f(x|\theta, \Omega)\pi(\theta|\Omega)}{f(x|\Omega)}, \quad (11.1)$$

where the prior predictive density, $f(x|\Omega)$, is given by

$$f(x|\Omega) = \int_{\Theta} f(x|\theta, \Omega)\pi(\theta|\Omega)d\theta, \quad (11.2)$$

The posterior density, calculated through Eq. (11.1), describes the uncertainty about θ after taking the data into account, that is, $\pi(\theta|x, \Omega)$ is the posterior inference about θ , according to which it is possible to appraise the variability of θ .

As implied by Eqs. (11.1) and (11.2), the set Ω is present in every step of calculation. Therefore, for the sake of simplicity, this symbol will be suppressed in forthcoming equations. Another relevant fact in this context concerns the denominator of Eq. (11.1), which is expanded in Eq. (11.2): since the integration of Eq. (11.2) is carried out over the whole parameter space, the prior predictive distribution is actually a constant, and as such, it has the role of normalizing the right-hand side of Eq. (11.1). Therein arises another fairly common way of representing Bayes' theorem, as written as

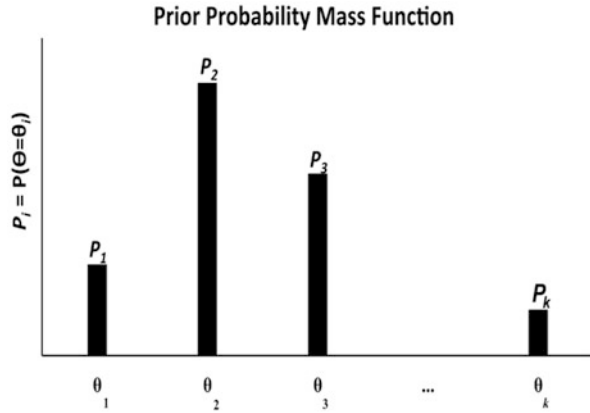
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta), \quad (11.3)$$

or, alternatively,

$$\text{posterior density} \propto \text{likelihood} \times \text{prior density}. \quad (11.4)$$

According to Ang and Tang (2007), Bayesian analysis is particularly suited for engineering problems, in which the available information is limited and often a subjective decision is required. In the case of parameter estimation, the engineer has, in some cases, some prior knowledge about the quantity on which inference is carried out. In general, it is possible to establish, with some degree of belief, which outcomes are more probable than others, even in the absence of any observational experiment concerning the variable of interest.

Fig. 11.1 Prior probability mass function of variable θ (adapt. Ang and Tang, 2007)



A hydrologist, for example, supported by his/her professional experience or having knowledge on the variation of past flood stages in river reaches neighboring a study site, even without direct monitoring, can make subjective preliminary evaluations (or elicitation) of the probability of the water level exceeding some threshold value. This can be a rather vague assessment, such as “not very likely” or “very likely,” or a more informative and quantified assessment derived from data observed at nearby gauging stations. Even such rather subjective information can provide important elements for the analysis and be considered as part of a logical and systematic analysis framework through Bayes’ theorem.

A simple demonstration of how Bayes’ theorem can be employed to update current expert knowledge makes use of discrete random variables. Assume that a given variable θ can only take values from the discrete set $\theta_i, i = 1, 2, \dots, k$, with respective probabilities $p_i = P(\theta = \theta_i)$. Assume further that after inferring the values of p_i , new information ε is gathered by some data collecting experiment. In such a case, the values of p_i should be updated in the light of the new information ε .

The values of p_i , prior to obtaining the new information ε , provide the prior distribution of θ , which is assumed to have already been elicited and summarized in the form of the mass function depicted in Fig. 11.1.

Equation (11.1), as applied to a discrete variable, may be rewritten as

$$P(\theta = \theta_i | \varepsilon) = \frac{P(\varepsilon | \theta = \theta_i)P(\theta = \theta_i)}{\sum_{i=1}^k P(\varepsilon | \theta = \theta_i)P(\theta = \theta_i)}, \quad i = 1, 2, \dots, k, \quad (11.5)$$

where,

- $P(\varepsilon | \theta = \theta_i)$ denotes the likelihood or conditional probability of observing ε , given that θ_i is true;

- $P(\theta = \theta_i)$ represents the prior mass of θ , that is, the knowledge about θ before ε is observed; and
- $P(\Theta = \theta_i | \varepsilon)$ is the posterior mass of θ , that is, the knowledge about θ after taking ε into account.

The denominator of Eq. (11.5) is the normalizing or proportionality constant, likewise to the previously mentioned general case.

The expected value of Θ can be a Bayesian estimator of θ , defined as

$$\hat{\theta} = E(\Theta | \varepsilon) = \sum_{i=1}^k \theta_i P(\Theta = \theta_i | \varepsilon) \quad i = 1, 2, \dots, k \quad (11.6)$$

Equation (11.6) shows that, unlike classical parameter estimation, both the observed data, taken into account via the likelihood function, and the prior information, be it subjective or not, are taken into account by the logical structure of Bayes' theorem. Example 11.1 illustrates these concepts.

Example 11.1 A large number of extreme events in a certain region may indicate the need for developing a warning system for floods and emergency plans against flooding. With the purpose of evaluating the severity of rainfall events over that region, suppose a meteorologist has classified the sub-hourly rainfall events with intensities of over 10 mm/h as extreme. Those with intensities lower than 1 mm/h were discarded. Assume that the annual proportion of extreme rainfall events as related to the total number of events can only take the discrete values $\theta = \{0.0, 0.25, 0.50, 0.75, \text{ and } 1.0\}$. This is, of course, a gross simplification, since a proportion can vary continuously between 0 and 1. Based on his/her knowledge of the regional climate, the meteorologist has evaluated the probabilities respectively associated with the proportions $\theta_i, i = 1, \dots, 5$, which are summarized in the chart of Fig. 11.2.

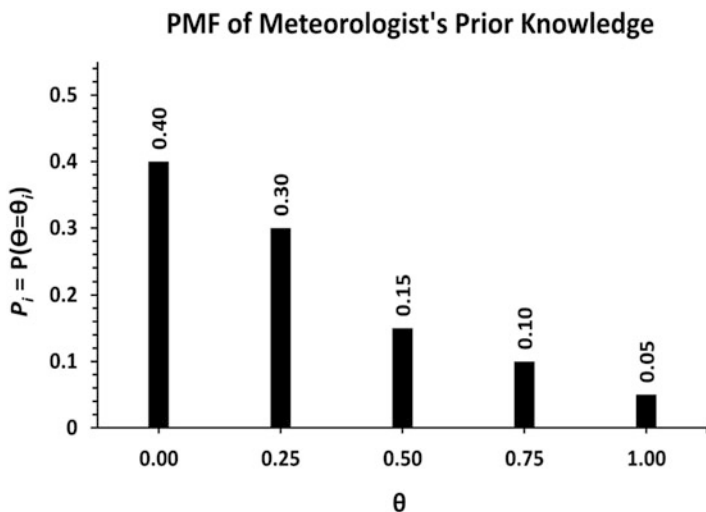


Fig. 11.2 Prior knowledge of the meteorologist on the probability mass function of the annual proportions of extreme rainfall events

Solution In the chart of Fig. 11.2, θ is the annual proportion of extreme rainfall events as related to the total number of events. For instance, the annual probability that none of the rainfall events is extreme is 0.40. Thus, based exclusively on the meteorologist's previous knowledge, the average ratio of events is given by

$$\begin{aligned}\hat{\theta}' &= E(\theta|\varepsilon) = 0.0 \times 0.40 + 0.25 \times 0.30 + 0.50 \times 0.15 + 0.75 \times 0.10 + 1.0 \times 0.05 \\ &= 0.275\end{aligned}$$

Suppose a rainfall gauging station has been installed at the site and that, 1 year after the start of rainfall monitoring, none of observed events could be classified as extreme. Then, the meteorologist can use the new information to update his/her prior belief using Bayes' theorem in the form of Eq. (11.5), or

$$P(\theta = 0.0|\varepsilon = 0.0) = \frac{P(\varepsilon=0.0|\theta=0.0)P(\theta=0.0)}{\sum_{i=1}^k P(\varepsilon=0.0|\theta=\theta_i)P(\theta=\theta_i)}, \text{ which, with the new data, yields}$$

$$\begin{aligned}P(\theta=0.0|\varepsilon=0.0) &= \frac{1.0 \times 0.40}{1.0 \times 0.40 + 0.75 \times 0.30 + 0.50 \times 0.15 + 0.25 \times 0.10 + 0.0 \times 0.05} \\ &= 0.552\end{aligned}$$

In the previous calculations, $P(\varepsilon = 0.0|\theta = \theta_i)$ refers to the probability that no extreme events happen, within a universe where $100\theta_i\%$ of events are extreme. Thus, $P(\varepsilon = 0.0|\theta = 0)$ refers to the probability of no extreme events happening, within a universe where 0% of the events are extreme, which, of course, is 100%. The remaining posterior probabilities are obtained in a likewise manner, as follows:

$$\begin{aligned}P(\theta=0.25|\varepsilon=0.0) &= \frac{0.75 \times 0.30}{1.0 \times 0.40 + 0.75 \times 0.30 + 0.50 \times 0.15 + 0.25 \times 0.10 + 0.0 \times 0.05} \\ &= 0.310\end{aligned}$$

$$\begin{aligned}P(\theta=0.50|\varepsilon=0.0) &= \frac{0.50 \times 0.15}{1.0 \times 0.40 + 0.75 \times 0.30 + 0.50 \times 0.15 + 0.25 \times 0.10 + 0.0 \times 0.05} \\ &= 0.103\end{aligned}$$

$$\begin{aligned}P(\theta=0.75|\varepsilon=0.0) &= \frac{0.25 \times 0.10}{1.0 \times 0.40 + 0.75 \times 0.30 + 0.50 \times 0.15 + 0.25 \times 0.10 + 0.0 \times 0.05} \\ &= 0.034\end{aligned}$$

$$\begin{aligned}P(\theta=1.00|\varepsilon=0.0) &= \frac{0.00 \times 0.05}{1.0 \times 0.40 + 0.75 \times 0.30 + 0.50 \times 0.15 + 0.25 \times 0.10 + 0.0 \times 0.05} \\ &= 0.000\end{aligned}$$

Figure 11.3 shows the comparison between the prior and the posterior mass functions. It is evident how the data, of 1 year of records among which no extreme

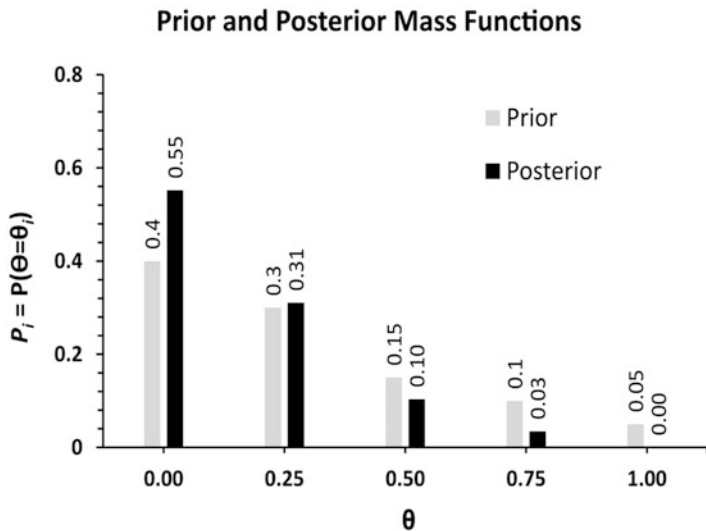


Fig. 11.3 Comparison between prior and posterior mass functions

event was observed, adjust the prior belief since the new evidence suggests that the expected proportion of extreme events is lower than initially thought.

The annual mean proportion of extreme events is given by:

$$\begin{aligned} \hat{\theta}'' = E(\theta|\epsilon) &= 0.0 \times 0.552 + 0.25 \times 0.310 + 0.50 \times 0.103 + 0.75 \times 0.034 + 1.0 \times 0.00 \\ &= 0.155 \end{aligned}$$

It should be noted that classical inference could hardly be used in this case, since the sample size is 1, which would result in $\hat{\theta} = 0$. Bayesian analysis, on the other hand, can be applied even when information is scarce. Suppose, now, that after a second year of rainfall gauging, the same behavior as the first year took place, that is, no extreme rainfall event was observed. This additional information can then be used to update the knowledge about θ through the same procedure described earlier. In such a case, the prior information for the year 2 is now the posterior mass function of year 1. Bayes' theorem can thus be used to progressively update estimates in light of newly acquired information. Figure 11.4 illustrates such a process of updating estimates, by hypothesizing the recurrence of the observed data, as in year 1 with $\epsilon = 0$, over the next 1, 5, and 10 years.

As shown in Fig. 11.4, after a few years of not a single occurrence of an extreme event, the evidence of no extreme events becomes stronger. As $n \rightarrow \infty$, the Bayesian estimation will converge to the frequentist one $[P(\epsilon = 0.0|\theta = 0) = 1]$. Example 11.1 illustrates the advantages of Bayesian inference. Nevertheless, it also reveals one of its major drawbacks: the subjectivity involved in eliciting prior

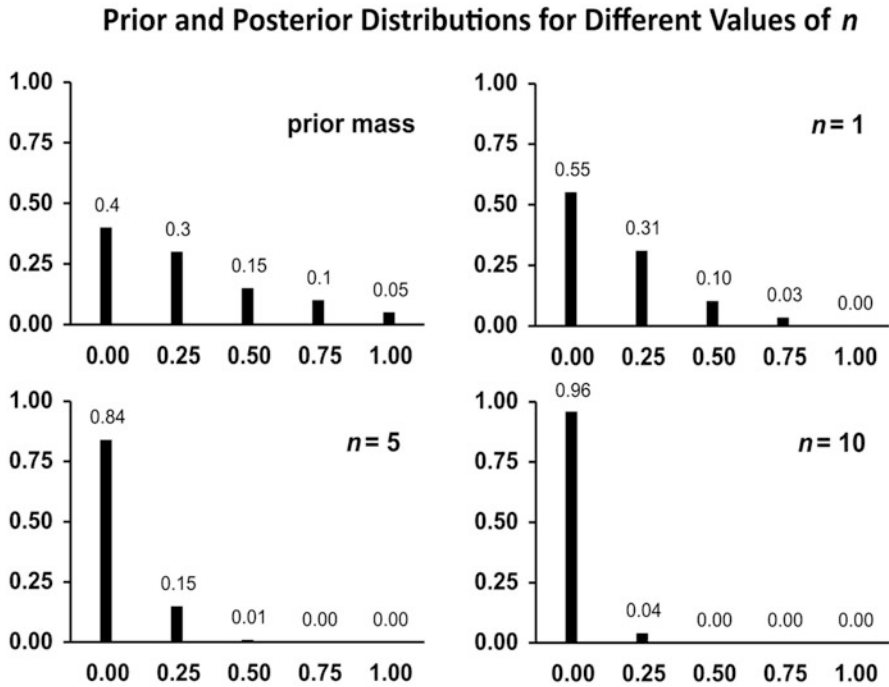


Fig. 11.4 Updating the prior mass functions after 1, 5, and 10 years in which no extreme rainfall event was observed

information. If another meteorologist had been consulted, there might have been a different prior probability proposal for Fig. 11.2, thus leading to different results. Therefore the decision will inevitably depend on how skilled or insightful is the source of previous knowledge elicited in the form of the prior distribution.

Bayesian inference does not necessarily entail subjectivity. Prior information may have sources other than expert judgement. Looking back at Example 11.1, the prior distribution of the ratio of extreme events can be objectively obtained through analysis of data from a rain gauging station. However, when the analyzed data are the basis for eliciting the prior distribution, they may not be used to calculate the likelihood function, i.e., each piece of information should contribute to only one of the terms of Bayes' theorem.

In Example 11.1 there was no mention of the parametric distribution of the variable under analysis. Nevertheless, in many practical situations it is possible to elicit a mathematical model to characterize the probability distribution of the quantity of interest. Such is the case in the recurrence time intervals of floods which are modeled by the geometric distribution (see Sect. 4.1.2), or the probability of occurrence of y floods with exceedance probability θ in N years, which is modeled by the binomial distribution (see Sect. 4.1.1). Bayes' theorem may be applied directly, although that requires other steps and different calculations than those presented so far, which depend on the chosen model. In the next paragraphs,

some of the steps required for a general application of Bayes' theorem are presented, taking the binomial distribution as a reference.

Assume Y is a binomial variate or, for short, $Y \sim B(N, \theta)$. Unlike the classical statistical analysis, the parameter θ is not considered as a fixed value but rather as a random variable that can be modeled by a probability distribution. Since θ can take values in the continuous range $[0, 1]$, it is plausible to assume that its distribution should be left- and right-bounded, and, as such, the Beta distribution is an appropriate candidate to model θ , i.e., $\theta \sim \text{Be}(a, b)$. Assuming that the prior distribution of θ is fully established (e.g., that the a and b values are given) and, having observed the event $Y = y$, Bayes' theorem provides the solution for the posterior distribution of θ as follows

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{\int_0^1 p(y|\theta)\pi(\theta)d\theta} \quad (11.7)$$

where

- $\pi(\theta|y)$ is the posterior density of θ after taking y into consideration;
- $p(y|\theta)$ is the likelihood of y for a given θ , that is, $\binom{N}{y}\theta^y(1-\theta)^{N-y}$;
- $\pi(\theta)$ is the prior density of θ , that is, the $\text{Be}(a, b)$ probability density function;
- $\int_0^1 p(y|\theta)\pi(\theta)d\theta$ is the normalizing or proportionality constant, hitherto represented by $p(y)$. It should be noted that $p(y)$ depends only upon y and is, thereby, a constant with respect to the parameter θ .

Under this setting, one can write

$$\pi(\theta|y) = \frac{\binom{N}{y}\theta^y(1-\theta)^{N-y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}}{p(y)} \quad (11.8)$$

After algebraic manipulation and grouping common terms, one obtains

$$\pi(\theta|y) = \frac{1}{p(y)}\binom{N}{y}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a+y-1}(1-\theta)^{b+N-y-1} \quad (11.9)$$

or

$$\pi(\theta|y) = c(N, y, a, b)\theta^{a+y-1}(1-\theta)^{b+N-y-1} \quad (11.10)$$

Note, in Eq. (11.10), that the term $\theta^{a+y-1}(1-\theta)^{b+N-y-1}$ is the kernel of the $\text{Be}(a+y, b+N-y)$ density function. Since the posterior density must integrate to 1 over its domain, the independent function $c(N, y, a, b)$ must be

$$c(N, y, a, b) = \frac{\Gamma(a + b + N)}{\Gamma(a + y)\Gamma(b + N - y)} \quad (11.11)$$

As opposed to Example 11.1, here it is possible to evaluate the posterior distribution analytically. Furthermore, any inference about θ , after taking the data point into account, can be carried out using $\pi(\theta|y)$. For example, the posterior mean is given by

$$E[\theta|y] = \frac{a + y}{a + b + N} \quad (11.12)$$

which can be rearranged as

$$E[\theta|y] = \frac{a + b}{a + b + N} \left(\frac{a}{a + b} \right) + \frac{N}{a + b + N} \left(\frac{y}{N} \right) \quad (11.13)$$

For the sake of clarity, Eq. (11.12) can be rewritten as

$$E[\theta|y] = \frac{a + b}{a + b + N} \times \{\text{prior mean of } \theta\} + \frac{n}{a + b + N} \times \{\text{data average}\} \quad (11.14)$$

Equation (11.13) shows that the posterior distribution is a balance between prior and observed information. As in Example 11.1, as the sample increases, prior information or prior knowledge becomes less relevant when estimating θ and inference results should converge to those obtained through the frequentist approach. In this case,

$$\lim_{N \rightarrow \infty} E[\theta|y] = \frac{y}{N} \quad (11.15)$$

Example 11.2 Consider a situation similar to that shown in Example 4.2. Suppose, then, that the probability p that the discharge Q_0 will be exceeded in any given year is uncertain. An engineer believes that p has mean 0.25 and variance 0.01. Note that, in Example 4.2, p was not considered to be a random variable. Furthermore, it is believed that p follows a Beta distribution, i.e., $p \sim \text{Be}(a, b)$. Parameters a and b may be estimated by the method of moments as $\hat{a} = \bar{p} \left(\frac{\bar{p}}{S_p^2} - 1 \right) = 4.4375$ and $\hat{b} = (1 - \bar{p}) \left(\frac{\bar{p}}{S_p^2} - 1 \right) = 13.3125$, where $\bar{p} = 0.25$ and $S_p^2 = 0.01$. In a sense, the variance of p , as denoted by S_p^2 , measures the degree of belief of the engineer as to the value of p . Assume that after 10 years of observations, no flow exceeding Q_0 was recorded. (a) What is the updated estimate of p in light of the new information? (b) What is the posterior probability that Q_0 will be exceeded twice in the next 5 years?

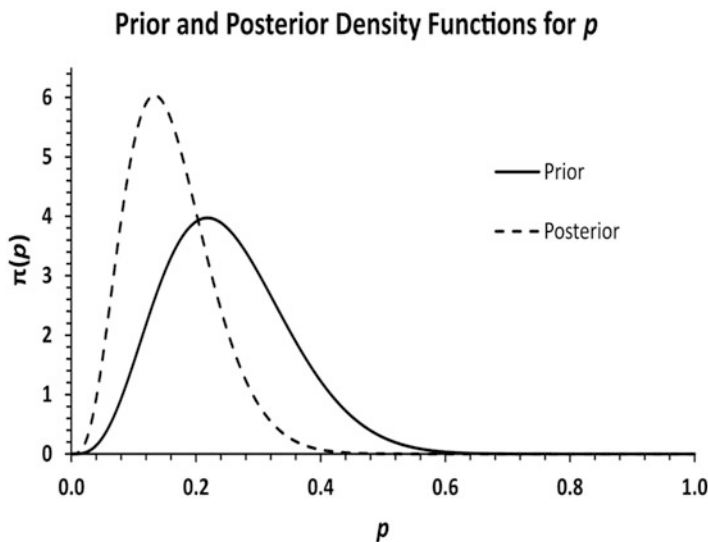


Fig. 11.5 Prior and posterior densities of p

Solution

- (a) As previously shown, $\pi(p|y) = \text{Be}(a + y, b + N - y) = \text{Be}(A, B)$, with $a = 4.4375$, $b = 13.3125$, $N = 10$ and $y = 0$. Equation (11.11) provides the posterior mean of p as $E[p|y] = \frac{4.4375+0}{4.4375+13.3125+10} = 0.1599$. Since no exceedances were observed in the 10 years of records, the posterior probability of flows in excess of Q_0 occurring is expected to decrease, or, in other words, there is empirical evidence that such events are more exceptional than initially thought. Figure 11.5 illustrates how the data update the distribution of p . Example 4.2 may be revisited using the posterior distribution for inference about the expected probability and the posterior credibility intervals, which are formally presented in Sect. 11.3.
- (b) The question here concerns what happens in the 5 next years, thus departing from the problem posed in Example 4.2. Recall that: (1) the engineer has a prior knowledge about p , which is formulated as $\pi(p)$; (2) no exceedance was recorded over a 10-year period, i.e., $y = 0$; (3) the information about p was updated to $\pi(p|y)$; and (4) he/she needs to evaluate the probability that a certain event \tilde{y} will occur in the next 5 years. Formally,

$$\begin{aligned}
 P(\tilde{Y} = \tilde{y} \mid Y = y) &= \int P(\tilde{Y} = \tilde{y}, Y = y) dp \\
 &= \int P(\tilde{Y} = \tilde{y}, p \mid Y = y) dp \\
 &= \int P(\tilde{Y} = \tilde{y} \mid p, Y) \pi(p \mid Y) dp \\
 &= \int \binom{N}{\tilde{y}} p^{\tilde{y}} (1 - p)^{N - \tilde{y}} f_{\text{Beta}(A, B)}(p) dp
 \end{aligned}$$

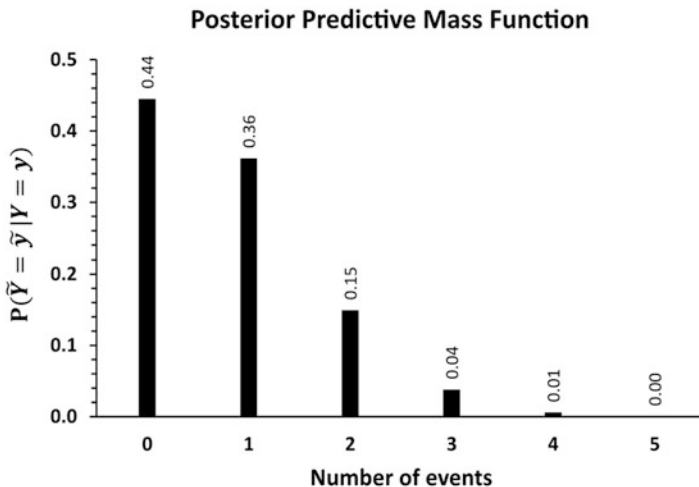


Fig. 11.6 Posterior predictive mass function of the number of events over 5 years

where $N = 5$, $\tilde{y} = 2$, $A = 4.4375$, and $B = 23.3125$. After algebraic manipulation one obtains

$$P(\tilde{Y} = \tilde{y} | Y = y) = \binom{N}{\tilde{y}} \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} \int p^{\tilde{y}} (1-p)^{N-\tilde{y}} p^{A+1} (1-p)^{B-1} dp.$$

Note that the integrand in this equation is the numerator of the PDF of the $Be(\tilde{y} + A, N - \tilde{y} + B)$ distribution. Since a density function must integrate to 1, over the domain of the variable, one must obtain

$$P(\tilde{Y} = \tilde{y} | Y = y) = \binom{N}{\tilde{y}} \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} \frac{\Gamma(\tilde{y}+A)\Gamma(N-\tilde{y}+B)}{\Gamma(N+A+B)} = 0.1494.$$

This is the posterior predictive estimate of $P(\tilde{y} = 2)$, since it results from the integration over all possible realizations of p . One could further define the probabilities associated with events $\tilde{y} = \{0, 1, 2, 3, 4, \dots, n\}$ over the next N years. Figure 11.6 illustrates the results for $N = 5$.

11.2 Prior Distributions

11.2.1 Conjugate Priors

In the examples discussed in the previous section, the product *likelihood* \times *prior density* benefited from an important characteristic: its mathematical form was such that, after some algebraic manipulation, a posterior density was obtained which belongs to the same family as the prior density (e.g., *Binomial* \times *Beta* \rightarrow *Beta*). Furthermore, in those cases, the proportionality constant (the denominator of Bayes' theorem) was obtained in an indirect way, without requiring integration.

This is the algebraic convenience of using *conjugate* priors, i.e., priors whose combination with a particular likelihood results in a posterior from the same family.

Having a posterior distribution with a known mathematical form facilitates statistical analysis and allows for a complete definition of the posterior behavior of the variable under analysis. However, as several authors point out (e.g., Paulino et al. 2003; Migon and Gamerman 1999) this property is limited to only a few particular combinations of models. As such, conjugate priors are not usually useful in most practical situations. Following is a non-exhaustive list of conjugate priors.

- Normal distribution (known standard deviation σ)

Notation: $X \sim N(\mu, \sigma)$

Prior: $\mu \sim N(\zeta, \tau)$

Posterior: $\mu|x \sim N(v(\sigma^2\zeta + \tau^2x), \tau\sigma\sqrt{v})$ with $v = \frac{1}{\sigma^2 + \tau^2}$

- Normal distribution (known mean μ)

Notation: $X \sim N(\mu, \sigma)$

Prior: $\sigma \sim \text{Ga}(\alpha, \beta)$

Posterior: $\sigma|x \sim \text{Ga}\left(\alpha + \frac{1}{2}, \beta + \frac{(\mu-x)^2}{2}\right)$

- Gamma distribution

Notation: $X \sim \text{Ga}(\theta, \eta)$

Prior: $\eta \sim \text{Ga}(\alpha, \beta)$

Posterior: $\eta|x \sim \text{Ga}(\alpha + \theta, \beta + x)$

- Poisson distribution

Notation: $Y \sim P(\nu)$

Prior: $\nu \sim \text{Ga}(\alpha, \beta)$

Posterior: $\nu|y \sim \text{Ga}(\alpha + y, \beta + 1)$

- Binomial distribution

Notation: $Y \sim B(N, p)$

Prior: $p \sim \text{Be}(\alpha, \beta)$

Posterior: $p|y \sim \text{Be}(\alpha + y, \beta + N - y)$

11.2.2 Non-informative Priors

In some situations there is a complete lack of prior knowledge about a given parameter. It is not straightforward to elicit a prior distribution that reflects total ignorance about such a parameter. In these cases, the so-called non-informative priors or vague priors can be used.

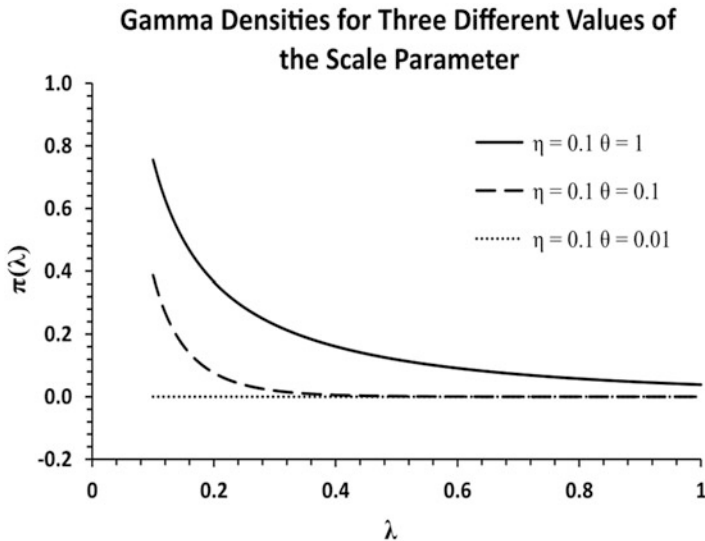


Fig. 11.7 Gamma densities for some values of the scale parameter

A natural impulse for modelers to convey non-information, in the Bayesian sense, is to attribute the same prior density to every possible value of the parameter. In that case, the prior must be a uniform distribution, that is $\pi(\theta) = k$. The problem with that formulation is that when θ has an unbounded domain, the prior distribution is improper, that is, $\int \pi(\theta) d\theta = \infty$. Although the use of proper distributions is not mandatory, it is considered to be a good practice in Bayesian analysis. Robert (2007) provides an in-depth discussion about the advantages of using proper prior distributions. A possible alternative to guarantee that $\int \pi(\theta) d\theta = 1$ is to use the so-called *vague* priors, which are parametric distributions with large variances such that they are, at least locally, nearly flat. Figure 11.7 illustrates this rationale for a hypothetical parameter λ , which is Gamma-distributed. Note how the Gamma density, with a very small scale parameter, is almost flat. Another option is to use a normal density with a large variance.

Another useful option is to use a Jeffreys prior. A Jeffreys prior distribution has a density defined as

$$\pi(\theta) \propto [I(\theta)]^{1/2} \tag{11.16}$$

where $I(\theta)$ denotes the so-called Fisher information about the parameter θ , as given by

$$I(\theta) = E \left[-\frac{\partial^2 \ln L(x|\theta)}{\partial \theta^2} \right] \quad (11.17)$$

and $L(x|\theta)$ represents the likelihood of x , conditioned on θ . Following is an example of application of a Jeffreys prior.

Example 11.3 Let $X \sim P(\theta)$. Specify the non-informative Jeffreys prior distribution for θ (adapted from Ehlers 2016).

Solution The log-likelihood function of the Poisson distribution can be written as

$\ln L(x|\theta) = -N\theta + \ln(\theta) \sum_{i=1}^N x_i - \ln \left(\prod_{i=1}^N x_i! \right)$ of which, the second-order derivative is

$$\frac{\partial^2 \ln L(x|\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left[-N + \frac{\sum_{i=1}^N x_i}{\theta} \right] = -\frac{\sum_{i=1}^N x_i}{\theta^2}. \text{ Then,}$$

$$I(\theta) = \frac{1}{\theta^2} E \left[\sum_{i=1}^N x_i \right] = \frac{N}{\theta} \propto \theta^{-1}.$$

Incidentally, that is the same prior density as the conjugate density of the Poisson model $\text{Ga}(\alpha, \beta)$, with $\alpha = 0.5$ and $\beta \rightarrow 0$. In general, with a correct specification of parameters, the conjugate prior holds the characteristics of a Jeffreys prior distribution.

11.2.3 Expert Knowledge

Although it is analytically convenient to use conjugate priors or non-informative priors, these solutions do not necessarily assist the modeler in incorporating any existing prior knowledge or belief into the analysis. In most cases, knowledge about a certain quantity does not exist in the form of a particular probabilistic model. Hence the expert must build a prior distribution from whatever input, be it partial or complete, that he/she has. This issue is crucial in Bayesian analysis and while there is no unique way of choosing a prior distribution, the procedures implemented in practice generally involve approximations and subjective determination. Robert (2007) explores in detail the theoretical and practical implications of the choice of prior distributions. In Sect. 11.5, a real-world application is described in detail, which includes some ideas about how to convert prior information into a prior distribution.

In the hydrological literature there are some examples of prior parameter distributions based on expert knowledge. Martins and Stedinger (2000) established the so-called “geophysical prior” for the shape parameter of the GEV distribution, based on past experience gained in previous frequency analyses of hydrologic maxima. The geophysical prior is given by a Beta density in the range $[-0.5, 0.5]$ and defined as

$$\pi(\kappa) = \frac{(0.5 + \kappa)^5 (0.5 - \kappa)^8}{B(6, 9)} \quad (11.18)$$

where $B(\cdot)$ is the Beta function. Another example is provided by Coles and Powell (1996) who proposed a method for eliciting a prior distribution for all GEV parameters based on a few “rough” quantile estimates made by expert hydrologists.

11.2.4 Priors Derived from Regional Information

While their “geophysical prior” was elicited based on their expert opinion about a statistically reasonable range of values taken by the GEV shape parameter, Martins and Stedinger (2000) advocate the pursuit of regional information from nearby sites to build a more informative prior for κ . In fact, as regional hydrologic information exists and, in cases, can be abundant, the Bayesian analysis framework provides a theoretically sound setup to formally include it in the statistical inference. There are many examples in the technical literature of prior distributions derived from regional information for frequency analysis of hydrological extremes (see Viglione et al. 2013 and references therein).

A recent example of such an approach is given in Silva et al. (2015). These authors used the Poisson-Pareto model (see Sect. 8.4.5), under a peaks-over-threshold approach, to analyze the frequency of floods in the Itajaí-açu River, in Southern Brazil, at the location of Apiúna. Since POT analysis is difficult to automate, given the subjectivity involved in the selection of the threshold and independent flood peaks, Silva et al. (2015) exploited the duality of the shape parameter of the Generalized Pareto (GPA) distribution for exceedance magnitudes and of the GEV distribution of annual maxima by extensively fitting the GEV distribution to 138 annual maximum flood samples in the region (the 3 southernmost states of Brazil), using maximum likelihood estimation and the resulting estimates of the shape parameter to construct a prior distribution for that parameter. Figure 11.8 shows the location of the 138 gauging stations and the spatial distribution of the estimated values of the GEV shape parameter.

A Normal distribution was fitted to the obtained estimates of κ . Figure 11.9 shows the histogram of the estimates of the GEV shape parameter and the PDF of the fitted Normal distribution. Figure 11.9 also shows, as a reference, the PDF of the geophysical prior elicited by Martins and Stedinger (2000). Silva et al. (2015) found that, while using the geophysical prior is worthwhile in cases where no other prior information about κ is available, in that particular region it did not adequately fit the data.

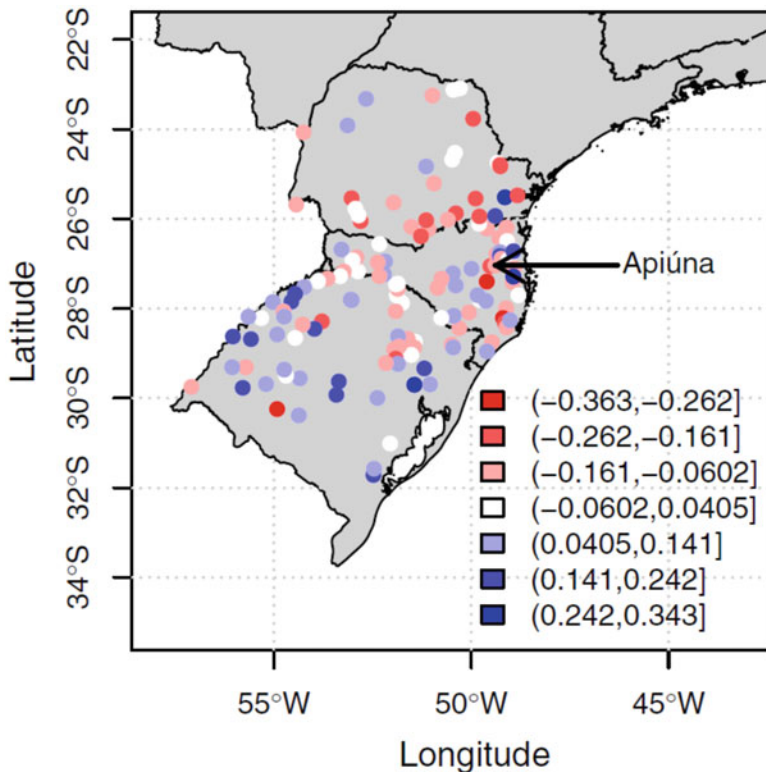


Fig. 11.8 Map of the south of Brazil with the location of the flow gauging stations used to elicit a prior distribution for the GEVshape parameter κ and spatial distribution of κ values (adapted from Silva et al. 2015)

11.3 Bayesian Estimation and Credibility Intervals

Bayesian estimation has its roots in *Decision Theory*. Bernardo and Smith (1994) argue that Bayesian estimation is fundamentally a decision problem. This section briefly explores some essential aspects of statistical estimation under a Bayesian approach. A more detailed description of such aspects can be found in Bernardo and Smith (1994) and Robert (2007).

Under the decision-oriented rationale for Bayesian estimation of a given parameter θ , one has a choice of a loss function, $\ell(\delta, \theta)$, which represents the loss or penalty due to accepting δ as an estimate of θ . The aim is to choose the estimator that minimizes the Bayes risk, denoted as BR and given by

$$BR = \iint \ell(\delta, \theta) f(x|\theta) \pi(\theta) dx d\theta \tag{11.19}$$

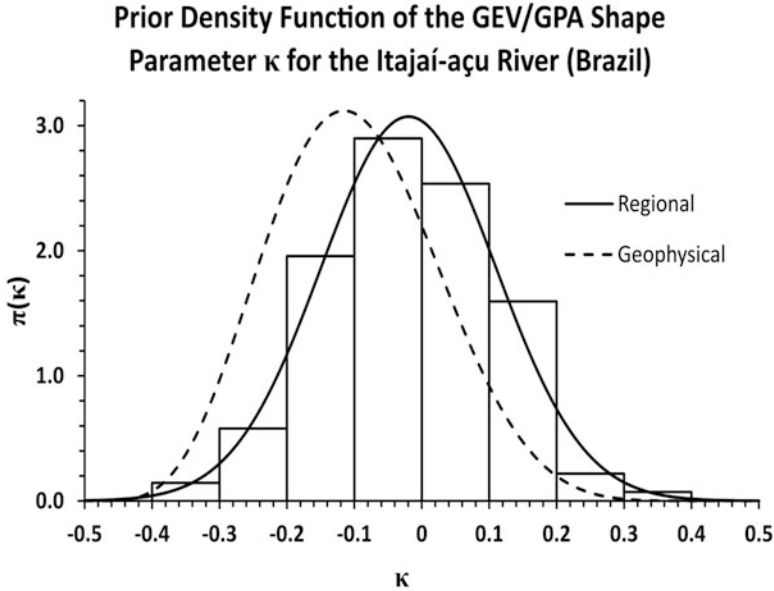


Fig. 11.9 Histogram and normal PDF fitted to the GEV shape parameter estimates, and the PDF of the geophysical prior as elicited by Martins and Stedinger (2000)

The inversion of the order of integration in Eq. (11.19), by virtue of Fubini’s theorem (see details in Robert 2007, p. 63), leads to the choice of the estimator δ which minimizes the posterior loss, that is the estimator δ_B of θ such that

$$\delta_B = \min_{\delta} E[\ell(\delta, \theta) | x] = \min_{\delta} \int_{\Theta} \ell(\delta, \theta) \pi(\theta | x) dx \tag{11.20}$$

The choice of the loss function is subjective and reflects the decision-maker’s judgment on the fair penalization for his/her decisions. According to Bernardo and Smith (1994), the main loss functions used in parameter estimation are:

- *Quadratic loss*, in which $\ell(\delta, \theta) = (\delta - \theta)^2$ and the corresponding Bayesian estimator is the posterior expectation $E[\pi(\theta | x)]$, provided that it exists;
- *Absolute error loss*, in which $\ell(\delta, \theta) = |\delta - \theta|$ and the Bayesian estimator is the posterior median, provided that it exists;
- *Zero-one loss*, in which $\ell(\delta, \theta) = \mathbf{1}_{\delta \neq \theta}$ in which $\mathbf{1}_{(a)}$ is the indicator function and the corresponding Bayesian estimator of θ is the posterior mode.

Robert and Casella (2004) point out two difficulties related to the calculation of δ . The first one is that, in general, the posterior density of θ , $\pi(\theta | x)$, does not have a closed analytical form. The second is that, in most cases, the integration of Eq. (11.20) cannot be done analytically.

Parameter estimation highlights an important distinction between the Bayesian and the frequentist approach to statistical inference: the way in which those two

approaches deal with uncertainties regarding the choice of the estimator. In frequentist analysis, this issue is addressed via the repeated sampling principle, and estimator performance based on a single sample is evaluated by the expected behavior of a hypothetical set of replicated samples collected under identical conditions, assuming that such replications are possible. The repeated sampling principle supports, for example, the construction of frequentist confidence intervals, CI (see Sect. 6.6). Under the repeated sampling principle, the confidence level $(1 - \alpha)$ of a CI is seen as the proportion of intervals, constructed on the basis of replicated samples under the exact same conditions as the available one, that contain the true value of the parameter.

The Bayesian paradigm, on the other hand, offers a more natural framework for uncertainty analysis by focusing on the probabilistic problem. In the Bayesian setting, the posterior variance of the parameter provides a direct measure of the uncertainty associated with its estimation. Credibility intervals (or posterior probability intervals) are the Bayesian analogues to the frequentist confidence intervals. The parameter θ , which is considered to be a random object, has $(1 - \alpha)$ posterior probability of being within the bounds of the $100(1 - \alpha)\%$ credibility interval. Thus the interpretation of the interval is more direct in the Bayesian case: there is a $(1 - \alpha)$ probability that the parameter lies within the bounds of the credibility interval. The bounds of credibility intervals are fixed and the parameter estimates are random, whereas in the frequentist approach the bounds of confidence intervals are random and the parameter is a fixed unknown value.

Credibility intervals can be built not only for a parameter but also for any scalar function of the parameters or any other random quantity. Let ω be any random quantity and $p(\omega)$ its probability density function. The $(1 - \alpha)$ credibility interval for ω is defined by the bounds (L, U) such that

$$\int_L^U p(\omega) d\omega = 1 - \alpha \quad (11.21)$$

Clearly, there is no unique solution for the credibility interval, even if $p(\omega)$ is unimodal. It is a common practice to adopt the highest probability density (HPD) interval, i.e., the interval $I \subseteq \Omega$, Ω being the domain of ω such that $P(\omega \in I) = 1 - \alpha$ and $p(\omega_1) \geq p(\omega_2)$ for every $\omega_1 \in I$ and $\omega_2 \notin I$ (Bernardo and Smith 1994). Hence the HPD interval is the narrowest interval such that $P(L \leq \omega \leq U) = 1 - \alpha$ and certainly provides a more natural and precise interpretation of probability statements concerning interval estimation of a random quantity.

11.4 Bayesian Calculations

The main difficulty in applying Bayesian methods is the calculation of the proportionality constant, or the prior predictive distribution, given by Eq. (11.2). To make inferences about probabilities, moments, quantiles, or credibility intervals, it is

necessary to calculate the expected value of any function of the parameter $h(\theta)$, as weighted by the posterior distribution of the parameter. Formally, one writes

$$E[h(\theta)|x] = \int_{\Theta} h(\theta)\pi(\theta|x)d\theta \quad (11.22)$$

The function h depends on the intended inference. For point estimation, h can be one of the loss functions presented in Sect. 11.3. In most hydrological applications, however, the object of inference is the prediction itself. If the intention is to formulate probability statements on the future values of the variable X , h can be the distribution of x_{n+1} given θ . In this case, one would have the posterior predictive distribution given by

$$F(x_{n+1}|x) = E[F(x_{n+1}|\theta)|x] = \int_{\Theta} F(x_{n+1}|\theta) \pi(\theta|x) d\theta \quad (11.23)$$

The posterior predictive distribution is, therefore, a convenient and direct way of integrating sampling uncertainties in quantile estimation.

The analytical calculation of integrals as in Eq. (11.22) is impossible in most practical situations, especially when the parameter space is multidimensional. However, numerical integration can be carried out using the Markov Chain Monte Carlo (MCMC) algorithms. According to Gilks et al. (1996), such algorithms allow for generating samples with a given probability distribution, such as $\pi(\theta|x)$, through a Markov chain whose limiting distribution is the target distribution. If one can generate a large sample from the posterior distribution of θ , say $\theta_1, \theta_2, \dots, \theta_m$, then the expectation given by Eq. (11.22) may be approximated by Monte Carlo integration. As such,

$$E[h(\theta)|x] \approx \frac{1}{m} \sum_{i=1}^m h(\theta_i) \quad (11.24)$$

In other terms, the population mean of h is estimated by the sample mean of the generated posterior sample.

When the sample $\{\theta_i\}$ is of IID variables, then, by the law of large numbers (see solution to Example 6.3), the approximation of the population mean can be as accurate as possible, requiring that the generated sample size m be increased (Gilks et al. 1996). However, obtaining samples of IID variables with density $\pi(\theta|x)$ is not a trivial task, as pointed out by Gilks et al. (1996), especially when $\pi(\theta|x)$ has a complex shape. In any case, the elements of $\{\theta_i\}$ need not be independent amongst themselves for the approximations to hold. In fact it is required only that the elements of $\{\theta_i\}$ be generated by a process which proportionally explores the whole support of $\pi(\theta|x)$.

To proceed, some definitions are needed. A Markov chain is a stochastic process $\{\theta_t, t \in T, \theta_t \in S\}$, where $T = \{1, 2, \dots\}$ and S represents the set of possible states of

θ , for which, the conditional distribution of θ_t at any t , given $\theta_{t-1}, \theta_{t-2}, \dots, \theta_0$, is the same as the distribution of θ_t , given θ_{t-1} , that is

$$P(\theta_{t+1} \in A | \theta_t, \theta_{t-1}, \dots, \theta_0) = P(\theta_{t+1} \in A | \theta_t), \quad A \subseteq S \quad (11.25)$$

In other terms, a Markov chain is a stochastic process in which the next state is dependent only upon the previous one. The Markov chains involved in MCMC calculations should generally have the following properties:

- *Irreducibility*, meaning that, regardless of its initial state, the chain is capable of reaching any other state in a finite number of iterations with a nonzero probability;
- *Aperiodicity*, meaning that the chain does not keep oscillating between a set of states in regular cycles; and
- *Recurrence*, meaning that, for every state I , the process beginning in I will return to that state with probability 1 in a finite number of iterations.

A Markov chain with the aforementioned characteristics is termed *ergodic*. The basic idea of the MCMC sampling algorithm is to obtain a sample with density $\pi(\theta|x)$ by building an ergodic Markov chain with: (1) the same set of states as θ ; (2) straightforward simulation; and, (3) the limiting density $\pi(\theta|x)$.

The Metropolis algorithm (Metropolis et al. 1953) is well-suited to generate chains with those requirements. That algorithm was developed in the Los Alamos National Laboratory, in the USA, with the objective of solving problems related to the energetic states of nuclear materials using the calculation capacity of early programmable computers, such as the MANIAC (Mathematical Analyzer, Numerical Integrator and Computer). Although the method gained notoriety in 1953 through the work of physicist Nicholas Metropolis (1915–1999) and his collaborators, the algorithm development had contributions from several other researchers, such as Stanislaw Ulam (1909–1984), John Von Neumann (1903–1957), Enrico Fermi (1901–1954), and Richard Feynman (1918–1988), among others. Metropolis himself admitted that the original ideas were due to Enrico Fermi and dated from 15 years before the date it was first published (Metropolis, 1987). Further details on the history of the development of the Metropolis algorithm can be found in Anderson (1986), Metropolis (1987) and Hitchcock (2003).

The Metropolis algorithm was further generalized by Hastings (1970) into the version widely known and used today. The algorithm uses a reference or jump distribution $g(\theta^*|\theta_t, x)$, from which it is easy to obtain samples of θ through the following steps:

The Metropolis–Hastings algorithm for generating a sample with density $\pi(\theta|x)$:

Initialize θ_0 ; $t \leftarrow 0$

Repeat {

Generate $\theta^* \sim g(\theta^*|\theta_t, x)$

Generate $u \sim \text{Uniform}(0,1)$

Calculate $\alpha_{\text{MH}}(\theta^*|\theta_t, x) = \min\left\{1, \frac{\pi(\theta^*|x) g(\theta_t|\theta^*, x)}{\pi(\theta_t|x) g(\theta^*|\theta_t, x)}\right\}$
 If $u \leq \alpha_{\text{MH}}(\theta^*|\theta_t, x)$
 $\theta_{t+1} \leftarrow \theta^*$
 Else
 $\theta_{t+1} \leftarrow \theta_t$
 $t \leftarrow (t+1)$
 }

An important aspect of the algorithm is that the acceptance rules are calculated using the ratios of posterior densities $\pi(\theta^*|x)/\pi(\theta_t|x)$, thus dismissing the calculation of the normalizing constant. The generalization proposed by Hastings (1970) concerns the properties of the jump distribution $g(\cdot|\cdot)$: the original algorithm, named random-walk Metropolis, required a symmetrical jump distribution, such that $g(\theta_i|\theta_j) = g(\theta_j|\theta_i)$. In this case the simpler acceptance rule is

$$\alpha_{\text{RWM}}(\theta^*|\theta_t, x) = \min\left\{1, \frac{\pi(\theta^*|x)}{\pi(\theta_t|x)}\right\} \quad (11.26)$$

Robert and Casella (2004) point out that after a large number of iterations, the resulting Markov chain may eventually reach equilibrium, after a sufficiently large number of iterations, i.e., its distribution converges to the target distribution. After convergence, all the resulting samples have the posterior density, and the expectations expressed by Eq. (11.22) may be approximated by Monte Carlo integration, with the desired precision.

As in most numerical methods, the MCMC samplers require some fine tuning. The choice of the jump distribution, in particular, is a key element for an efficient application of the algorithm. As Gilks et al. (1996) argue, in theory, the Markov chain will converge to its limiting distribution regardless of which jump distribution is specified. However, the numerical efficiency of the algorithm, as conveyed by its convergence rate, is greatly determined by how the jump density g relates to the target density π : convergence will be faster if g and π are similar. Moreover, even when the chain reaches equilibrium, the way it explores the support of the target distribution might be slow, thus requiring a large number of iterations. From the computational perspective, g should be a density that proves to be practical to evaluate at any point and to sample from. Furthermore, it is a good practice to choose a jump distribution with heavier tails than the target, in order to insure an adequate exploration of the support of $\pi(\theta|x)$. Further details on how to choose an adequate jump distribution are given by Gilks et al. (1996).

Another important aspect of adequate MCMC sampling is the choice of starting values: poorly chosen starting values might hinder the convergence of the chains for the first several hundreds or thousands of iterations. This can be controlled by discarding a sufficient number of iterations from the beginning of the chain, which is referred to as the *burn-in* period.

Currently, there are a large number of MCMC algorithms in the technical literature, all of them being special cases of the Metropolis–Hastings algorithm. One of the most popular algorithms is the Gibbs sampler, which is useful for multivariate analyses, and uses the complete conditional posterior distributions as the jump distribution. As a comprehensive exploration of MCMC algorithms is beyond the scope of this chapter, the following references are recommended for further reading: Gilks et al. (1996), Liu (2001), Robert and Casella (2004), and Gamerman and Lopes (2006).

Example 11.4 In order to demonstrate the use of the numerical methods discussed in this section, Example 11.2 is revisited with the random-walk Metropolis algorithm (acceptance rule given by Eq. 11.26). The following R code was used to generate a large sample from the posterior distribution of parameter p :

```
# prior density
pr <- function(theta) dbeta(theta, shape1=4.4375, shape2=13.3125)

# Likelihood function
ll <- function(theta) dbinom(0,10,theta)

# Unnormalized posterior density (defined for theta between 0 and 1)
unp <- function(theta) ifelse(theta >= 0 && theta <= 1, pr(theta)*ll(theta), 0)

theta_chain <- rep(NA, 100000)
theta_chain[1] <- 0.15 #Initialize

# Random-walk Metropolis algorithm
set.seed(123)
for (i in 1:99999) {
  # Jump or proposal density is Normal with standard deviation 0.05
  proposal <- rnorm(1, mean=theta_chain[i], sd=0.05)
  # Acceptance rule
  U <- runif(1)
  AR <- min(c(1, (unp(proposal)/unp(theta_chain[i]))))
  if (U <= AR) {
    theta_chain[i+1] <- proposal
  } else {
    theta_chain[i+1] <- theta_chain[i]
  }
}
```

The generated sample values are stored in the vector `theta_chain`. These values can be used for inferring any scalar function of the parameter. Figure 11.10 depicts the trace plot of the chain generated with the given code. A trace plot is a

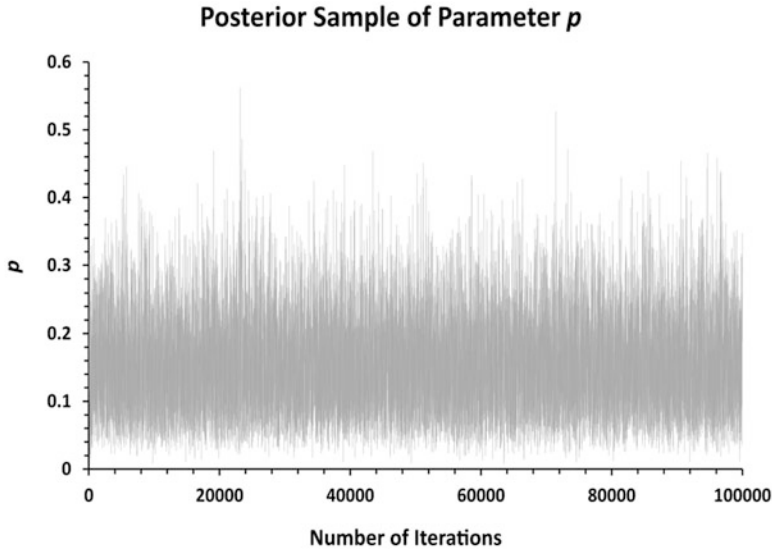


Fig. 11.10 Posterior sample of parameter p generated by the random-walk Metropolis algorithm

basic graphical tool used to detect clear signs of deviant or nonstationary behavior of generated chains, which may indicate a failure of convergence. It can also be used to determine the burn-in period when the starting values are poorly chosen. Visual inspection of a trace plot of a chain with good properties should not detect upward or downward trends or other nonstationary behavior, such as the chain getting stuck in certain regions of the parameter space. It should appear that each element of the plot is randomly sampled from the same target distribution. Figure 11.10, therefore, exemplifies a “good” traceplot.

The sample mean of the chain provides a point estimate for p , $E[p|y] = \frac{1}{N} \sum_{i=1}^N p_i = 0.1581$. This estimate is very close to the exact value of p , obtained in Example 11.2, $E[p|y] = 0.1599$. Figure 11.11 compares the exact solution of the posterior PDF of Example 11.2 with the histogram of the MCMC chain. It is clear that two solutions are very similar.

Even in the absence of a thorough monitoring of the MCMC chain convergence (see related procedures in Gilks et al. 1996; Liu 2001), Example 11.4 illustrates the capabilities of MCMC algorithms. In this case, it should be stressed that the posterior sample of the parameter was generated using the un-normalized posterior density, thereby dismissing the calculation of the constant of proportionality in the denominator of Bayes’ rule, which, in most practical situations, would reveal impossible to be made.

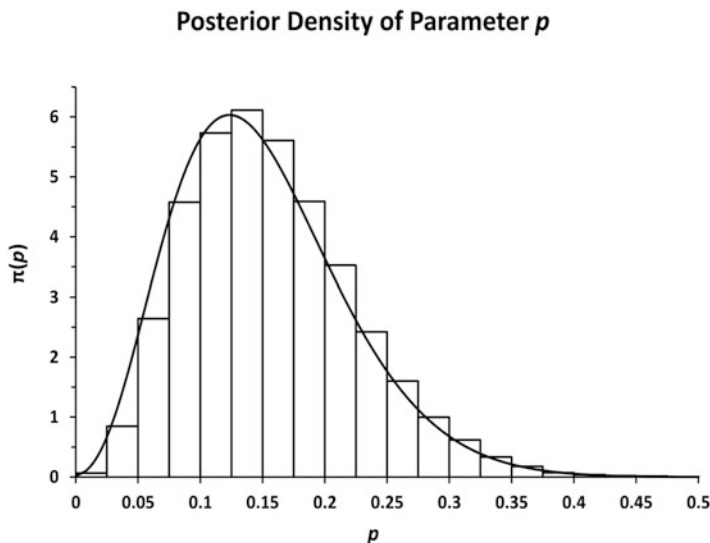


Fig. 11.11 Posterior density of parameter p . Exact solution (*continuous line*) and MCMC chain histogram

11.5 Example Application

This section presents an application of the principles of Bayesian analysis to a hydrology-related case study. The research was carried out by Fernandes et al. (2010). Readers interested in details of the case study are referred to this reference. The object of the study is related to the estimation of very rare extreme flood quantiles, with exceedance probabilities ranging from 10^{-6} to 2×10^{-3} , usually required for designing critical hydraulic structures, such as spillways of large dams with high potential flood hazards. The uncertainties associated with such estimates are admittedly very large and are not precisely quantifiable by standard procedures of statistical inference as the available samples of annual maximum floods have typical lengths ranging from 25 to 80 years. To tackle this problem, hydrologists often choose to estimate an upper bound for annual maximum floods, based on current knowledge of hydrological processes under extreme conditions. In this context, the concept of Probable Maximum Flood (PMF) is commonly used in connection to the design of major hydraulic structures (USNRC 1977; ICOLD 1987; FEMA 2004).

In short, the PMF is the upper bound of potential flooding in a given river section, resulting from a hypothetical rain storm with a critical depth and duration, deemed probable maximum precipitation (PMP), which should be preceded by very severe, but physically plausible hydrological and hydrometeorological conditions (see Sect. 8.3.2). The PMP, in turn, is formally defined by the World Meteorological Organization (WMO 1986) as “the greatest depth of precipitation for a given

duration meteorologically possible for a given size storm area at a particular location at a particular time of year, with no allowance made for long-time climatic trends.”

Although dam safety guidelines consider “quasi-deterministic” floods, such as the PMF, as a standard design criterion for large hydraulic structures, the estimation of a credible exceedance probability associated with such an extreme flood, in a way that risk-based decisions can be made, is not a trivial task (Dawdy and Lettenmaier 1987; Dubler and Grigg 1996; USBR 2004). Some conceptual changes on how frequency analysis is usually conducted are required, before associating an exceedance probability with the PMF estimate for a given catchment.

There are two major obstacles for merging the concepts of PMF and flood frequency analysis. Firstly, the available information of flow extremes is usually scarce since the available annual maximum samples usually span a few decades and very rarely over more than a century. As such, extrapolation of frequency curves for very rare quantiles, well beyond a credible range of extrapolation, is required, with all the uncertainty it entails (see Sect. 8.3). The second obstacle is related to the fact that many probability distributions used in flood frequency analysis have no upper bound and, therefore, do not accommodate the inherent concept of the PMF. Fernandes et al. (2010) propose the following steps of a workaround procedure:

- Adopting an upper-bounded probability distribution: although the use of such distributions is uncommon and even controversial among hydrologists, their structure is concurrent with limited extreme flood generating physical conditions, that is, they accommodate the notion of an upper bound for a flood in a given catchment.
- Analysis of paleohydrologic proxy data: as discussed in Sect. 8.2.3 and later on in this section, paleoflood data allow for a more comprehensive characterization of rare floods.
- Using the PMF for estimating the upper bound: in the application example described herein, the PMF is not used as a deterministic upper bound but rather informs the elicitation of a prior distribution for that upper bound.
- Using Bayesian inference methods: the Bayesian framework of analysis offers the means to aggregate the information from various sources.

The application example itself is presented in Sect. 11.5.3. The preceding Sect. 11.5.1 and 11.5.2 are necessary for presenting essential concepts and the formalism required to grasp Sect. 11.5.3.

11.5.1 Nonsystematic Flood Data

In recent decades there have been attempts to overcome the lack of data on extreme floods in streamflow records by including proxy information, deemed nonsystematic data, in flood frequency analysis (Stedinger and Cohn 1986; Francés et al. 1994; Francés 2001; Naulet 2002; Viglione et al. 2013). There are two main

sources of nonsystematic flood information: historical information, which refers to flood events directly observed or otherwise recorded by humans; and the so-called paleofloods, which correspond to floods that have occurred sometime in the Holocene epoch (approximately in the last 10,000 years) and can be reconstructed using remaining geological and/or botanical physical evidence.

Historical and paleohydrological information generally correspond to censored data which can be either upper-bounded (UB), or lower-bounded (LB), or double-bounded (DB). The type of censoring is determined by detectable limits of water levels of a particular flood during a given span of time. In any case, the exact maximum water level (or discharge) is not known with precision. For UB information, it is known that no flood has ever exceeded a certain level during the considered time span. For LB information, it is known that the flood has exceeded a given level. Finally for DB information, it is known that the maximum flood level is contained within an interval defined by two known bounds (LB,UB). Given some hypotheses, nonsystematic flood data can be incorporated into flood frequency analysis using appropriate statistical methods, thus potentially increasing the reliability of estimated extreme flood quantiles (Francés 2001).

11.5.2 Upper-Bounded Distributions

In the set of probability distributions that are commonly used by hydrologists in flood frequency analysis, some may be upper-bounded, depending on the particular combinations of the numerical values of their parameters. Some of those distributions may have up to 4 or 5 parameters (see Sect. 5.9). Others, such as the generalized extreme value (GEV) distribution is upper-bounded if its shape parameter is positive, which occurs when its skewness coefficient is less than 1.1396. Likewise, the log-Pearson type III (LP3) distribution has an upper limit if its skewness coefficient is negative.

Other upper-bounded probability models that should be mentioned are the Kappa (Hosking and Wallis 1997) and Wakeby distributions. The four-parameter Kappa model may be upper- or lower-bounded for a particular set of values of both its parameters (see Sect. 5.9.1). Analogously, the five-parameter Wakeby model may be upper-bounded for a particular combination of its three shape parameters (see Sect. 5.9.2). These are general distributional forms which can accommodate a number of bounded or unbounded models. For the application described herein, there is the additional requirement that the upper-bound should have an explicit parametric form. Fernandes et al. (2010) proposed the use of the four-parameter lognormal model (LN4) to explicitly incorporate the information provided by PMF estimates into frequency analysis.

The LN4 model branches from the following transformation

$$Y = \ln\left(\frac{X - \varepsilon}{\alpha - X}\right) \tag{11.27}$$

where $\varepsilon \in \mathfrak{R}_+$ denotes the lower bound of X , $\alpha \in \mathfrak{R}_+$ denotes the upper bound of X , and $Y \sim N(\mu_Y, \sigma_Y)$. Takara and Tosa (1999) point out that the LN4 model is not strongly affected by assuming its lower bound is zero. Taking advantage of that observation, the lower bound here is considered as $\varepsilon = 0$, thus rendering the model more parsimonious (one less parameter to estimate and one less prior distribution to elicit).

If $\varepsilon = 0$, the PDF of the variate $X \sim LN4(\mu_Y, \sigma_Y, \alpha)$ is given by

$$f_X(x|\theta) = \frac{\alpha}{x(\alpha - x)\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_Y^2}\left[\ln\left(\frac{x}{\alpha - X}\right) - \mu_Y\right]^2\right\} \tag{11.28}$$

with $0 \leq x \leq \alpha$, and the corresponding CDF is given by

$$F_X(x|\theta) = \Phi\left[\frac{1}{\sigma_Y}\ln\left(\frac{x}{\alpha - x}\right) - \frac{\mu_Y}{\sigma_Y}\right] \tag{11.29}$$

where Φ denotes the CDF of the standard Normal distribution.

Finally, considering the PDF given by Eq. (11.28) and a set of systematic and nonsystematic data, the likelihood function can be constructed in the manner described as follows. Let X_1, \dots, X_{Nex} be the sample, of size Nex , of annual maximum floods (systematic data). Upper-bounded censored data are denoted as UB_1, \dots, UB_{Nub} , lower-bounded censored data as LB_1, \dots, LB_{Nlb} , whereas DB_1, \dots, DB_{Ndb} denote the censored data which is double-bounded within the interval (LB, UB) . If the data are IID, then, the likelihood function is give by

$$\begin{aligned} L(x|\theta) &= \prod_i^{Nex} f_X(EX_i|\theta) \times \\ &\quad \prod_i^{Nub} F_X(UB_i|\theta) \times \\ &\quad \prod_i^{Nlb} [1 - F_X(LB_i|\theta)] \times \\ &\quad \prod_i^{Ndb} [F_X(UR_i|\theta) - F_X(LR_i|\theta)] \end{aligned} \tag{11.30}$$

11.5.3 Study Site and Data

The application example described here refers to the American River at Folsom Lake, located in the American State of California. This dam site was chosen particularly due to the availability of systematic and nonsystematic flood data. Figure 11.12 shows the schematic location of the study site.

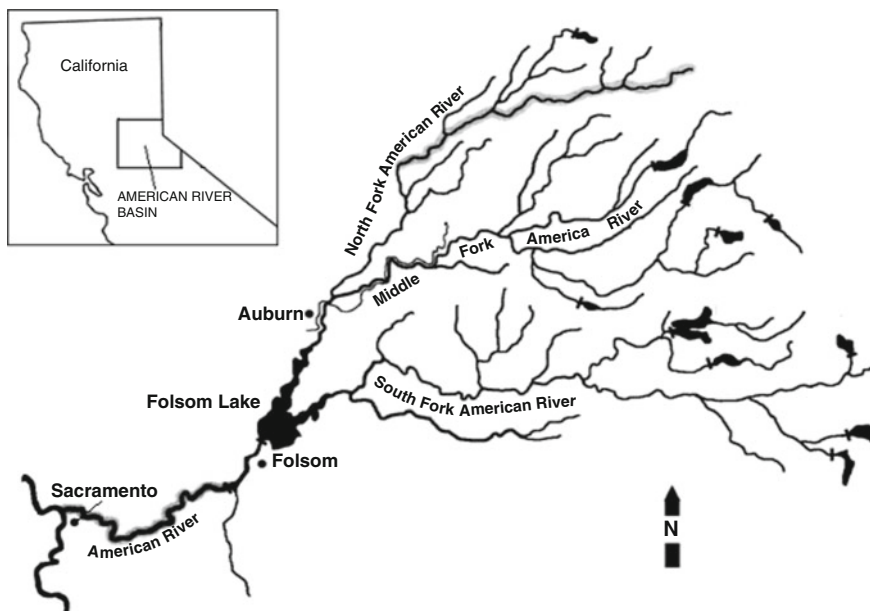


Fig. 11.12 American River basin (adapted from USBR 2002)

According to USBR (2002), the catchment of the American River at Folsom Lake has a drainage area of 4820 km² and its flows have been monitored by the US Geological Service since 1905. There are 52 years of systematic data. The annual maximum series passed the tests for randomness, independence and homogeneity and stationarity at the significance level of 5%, according to the nonparametric hypothesis tests presented in Sect. 7.4.

Regarding the nonsystematic data, studies conducted by USBR (2002) identified the occurrence of 4 distinct paleoflood levels dating back 2000 years before present (BP), all of them being UB censored. Furthermore, there are 5 DB censored floods. The chart in Fig. 11.13 illustrates the data sets used in the case study.

11.5.4 Prior Distribution of Parameters of the LN4 Model

In the Bayesian paradigm, the prior distribution is the mathematical synthesis of the degree of knowledge or belief that some expert has about the quantity of interest. The specification of the distribution is based on personal belief gathered through observation, or experience gained from similar situations, literature review, etc. The prior distribution is elicited before the data are observed.

Amongst the parameters of the LN4, the upper-bound α is perhaps the only one that shows a clear connection to climate and hydrological characteristics of the watershed under extreme hydrometeorological conditions. The hydrological

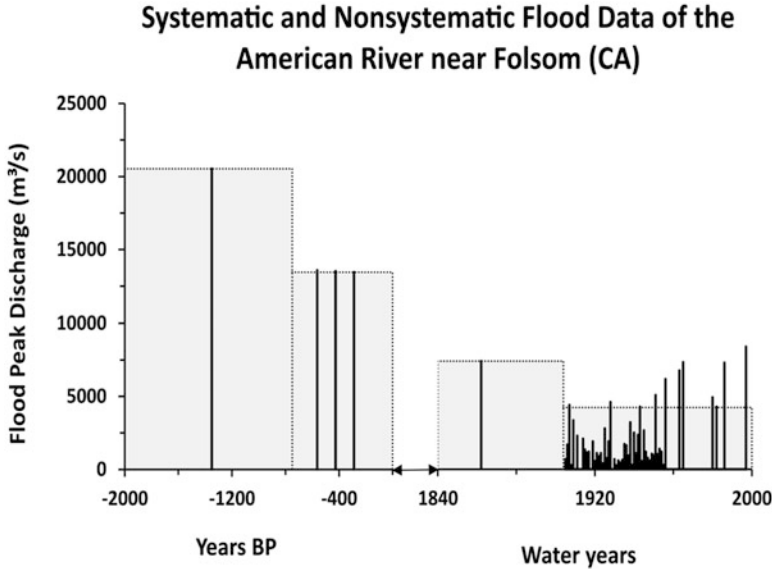


Fig. 11.13 Systematic and nonsystematic data of the American River near Folsom

interpretation of parameters μ_Y and σ_Y (respectively, the mean and standard deviation of the transformed LN4 variate Y) is more complex since it is not straightforward to identify any clear relationship between those parameters and physical characteristics of the catchment. Therefore, eliciting a prior distribution for these parameters is not simple, at least using this approach. A Normal distribution with a large variance was elicited for μ_Y , since it can take any real value, that is $\mu_Y \sim N(1.0, 10^{-6})$. Analogously, since σ_Y can only take positive values, the Gamma distribution $\sigma_Y \sim N(1.0, 10^{-8})$ was adopted as a prior for this parameter. Since these priors are proper, than the posterior parameter distributions should also be proper. Furthermore, it should be mentioned that the parametrization of the Normal and Gamma densities used in this example differ slightly from the parametrizations used elsewhere in this book. As such,

$$f_{\text{Normal}}(z|a, b) = \sqrt{\frac{b}{2\pi}} \exp\left[-\frac{b}{2}(z - a)^2\right] \text{ and } f_{\text{Gamma}}(z|a, b) = \frac{b^a z^{a-1}}{\Gamma(a)} \exp(-bz)$$

Unlike the location and scale parameters, the upper bound parameter is directly related to hydrometeorological phenomena in the catchment, thus allowing for a more informative elicitation of a prior distribution. In theory, if there were a set of PMF estimates for the catchment, acquired using the same methodology applied at different points in time, such data might have been used for eliciting a prior distribution for the upper bound parameter. Since that notional data set cannot exist, prior elicitation for that parameter must take a different approach based on

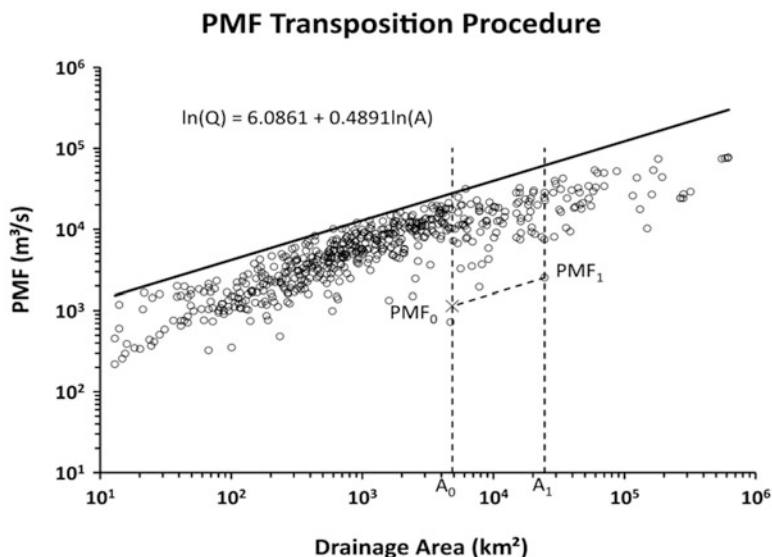


Fig. 11.14 Schematic of the PMF transposition procedure

available information. Fernandes et al. (2010) propose two methods based on a large set of PMF estimates for different North American catchments. The following is the detailed presentation of one of such methods.

The proposed method is based on the transposition of PMF estimates from other catchments to the catchment of interest. A data set of 561 PMF estimates compiled by the US Nuclear Regulatory Commission (USNRC 1977) was used, referring to catchments of varying sizes and characteristics scattered throughout the territory of the USA. Figure 11.14 shows the applied procedure: rather than transposing the PMF estimates directly to the study site, the envelope PMF curve was used, as proposed by USNRC (1977), where A_0 is the drainage area of the American River at Folsom Lake, resulting in transposing 561 PMF estimates to the catchment area of the study site, whose frequency analysis can inform a prior distribution for the upper bound of the LN4 distribution.

Fernandes (2009) showed that the Gamma distribution is a good candidate for modeling uncertainties related to the upper bound. Therefore, taking as reference the 561 PMF estimates transposed to the American River at Folsom Lake, and using the conventional method of moments, the prior distribution $\alpha \sim \text{Ga}(5.2, 0.00043)$ was elicited for the upper bound.

11.5.5 Posterior Distributions and Further Results

The posterior density of parameters is given by Eq. (11.1). Clearly there is no analytical solution for such a case, since it would require integration of the product

of the likelihood by the priors over the whole parametric space. For that reason, MCMC algorithms are applied, as discussed in Sect. 11.4. The following questions are posed before an MCMC algorithm is applied:

- What jump distribution should be used?
- Which sampling algorithm should be applied (Metropolis–Hastings, Gibbs sampler, slice sampler...)?
- How does one check whether the Markov chain has converged to the target distribution?

There are no straightforward answers to these questions. It is up to the practitioner to fine-tune the samplers, try different algorithms and jump distributions until a method is found that produces well-converged chains so that usual convergence theorems from Markov theory apply. Fortunately, there are several freely available software packages that provide useful tools for tuning and evaluating MCMC samplers. A thorough presentation of such packages is beyond the scope of this chapter. Gilks et al. (1996) and Albert (2009) are useful references on MCMC software.

In the application described herein, the WinBUGS software was used (Lunn et al. 2000). WinBUGS is a free user-friendly tool that requires minimal programming skills and is available from <http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/> (accessed: 21st February 2016). The following code was used at different stages of analysis of the presented case study:

```
# model with systematic and nonsystematic data
model {
    # Likelihood of the LN4 distribution
    #SYSTEMATIC DATA
    for (i in 1:NEX) {
        Z[i] <- abs(x[i]/(alpha-x[i]))
        index[i] <- step(alpha-x[i]) + 1
        L[i,1] <- 0
        L[i,2] <- (0.3989*alpha/sigma)*(1/ (x[i]*(alpha-x[i])))
    }
    *exp(-0.5*pow((1/sigma)*log(Z[i])-mi/sigma,2))
    #DB DATA
    for (i in NEX+1:NDB+NEX) {
        index[i] <- step(alpha-DBU[i-NEX]) + 1
        L[i,1] <- 0
        Zu[i-NEX] <- abs(DBU[i-NEX]/(alpha-DBU[i-NEX]))
        Zl[i-NEX] <- abs(DBL[i-NEX]/(alpha-DBL[i-NEX]))
        L[i,2] <- phi((1/sigma)*log(Zu[i-NEX])-mi/sigma)-phi
        ((1/sigma)*log(Zl[i-NEX])-mi/sigma)
    }
    #UB DATA - Level 1
    for (i in NDB+NEX+1:NDB+NEX+NYH[1]) {
        index[i] <- step(alpha-YH[1]) + 1
        L[i,1] <- 0
    }
}
```



```

    Zub1[i-NDB-NEX] <- abs(YH[1]/(alpha-YH[1]))
    L[i,2] <- phi((1/sigma)*log(Zub1[i-NDB-NEX])-mi/sigma)
  }
  #UB DATA - Level 2
  for (i in NDB+NEX+NYH[1]+1:NDB+NEX+NYH[1]+NYH[2]) {
    index[i] <- step(alpha-YH[2])+1
    L[i,1] <- 0
    Zub2[i-NDB-NEX-NYH[1]] <- abs(YH[2]/(alpha-YH[2]))
    L[i,2] <- phi((1/sigma)*log(Zub2[i-NDB-NEX-NYH[1]])-mi/
sigma)
  }
  #UB DATA- Level 3
  for (i in NDB+NEX+NYH[1]+NYH[2]+1:NDB+NEX+NYH[1]+NYH
[2]+NYH[3]) {
    index[i] <- step(alpha-YH[3])+1
    L[i,1] <- 0
    Zub3[i-NDB-NEX-NYH[1]-NYH[2]] <- abs(YH[3]/(alpha-YH
[3]))
    L[i,2] <- phi((1/sigma)*log(Zub3[i-NDB-NEX-NYH[1]-NYH
[2]])-mi/sigma)
  }
  #UB DATA- Level 4
  for (i in NDB+NEX+NYH[1]+NYH[2]+NYH[3]+1:NDB+NEX+NYH
[1]+NYH[2]+NYH[3]+NYH[4])
{
    index[i] <- step(alpha-YH[4])+1
    L[i,1] <- 0
    Zub4[i-NDB-NEX-NYH[1]-NYH[2]-NYH[3]] <- abs(YH
[4]/(alpha-YH[4]))
    L[i,2] <- phi((1/sigma)*log(Zub4[i-NDB-NEX-NYH[1]-NYH
[2]-NYH[3]])-mi/sigma)
  }
  for (i in 1:NDB+NEX+NYH[1]+NYH[2]+NYH[3]+NYH[4]) {
    dummy[i] <- 0
    dummy[i] ~ dgeneric(phi[i])
    phi[i] <- log(L[i,index[i]])
  }
  # PRIOR DENSITIES
  sigma ~ dgamma(1.0,1.0E-8)
  mi ~ dnorm(1.0,1.0E-6)
  alpha ~ dgamma(5.2,0.00043)
}

```

```

#Systematic and nonsystematic data
list(x=c(685,1691,4417,292,3370,2302,2098,1356,1152,1198,
        1911,569,1110,895,1104,396,2818,776,1917,4616,
        691,280,597,467,640,1724,1651,934,3228,309,
        2526,1099,2356,4304,569,2673,1195,790,595,1062,
        974,5097,1053,1407,1206,6201,6796,7362,4955,4304,
        7334,8438),
      NEX=52,
      DBL=c(7419,11327,11327,11327,16990),
      DBU=c(8495,15574,15574,15574,24069),
      NDB=5,
      YH=c(4248,7447,13451,20530),
      NYH=c(44,56,544,1299)
)

#Initial values
list(alpha=25000,
      sigma=2,
      mi=2)

```

By applying the WinBUGS code above, a chain of length 600,000 was generated. After a visual analysis of the trace plots of the chain, it was verified that it became stationary after the burn-in period of 100,000 realizations, which were discarded before proceeding with the analysis. Furthermore, the chains were thinned by retaining every 10th value, in order to remove autocorrelation. These actions resulted in a sample of 50,000 posterior parameters. Figure 11.15 shows the prior and posterior densities of the upper bound.

Figure 11.15 shows a clear disparity regarding the lower tails of the prior and posterior distributions. Prior to looking at the data there is no evidence regarding how low the upper bound could possibly be, which provides support for a prior distribution developing throughout the set of all positive real numbers. On the other hand, after taking the data into account, it does not seem logical to say that the upper bound can be lower than the maximum observed flood, so the support was modified for the posterior distribution. Table 11.1 shows some statistics for the prior and posterior distributions, illustrating how the systematic and nonsystematic data significantly reduced the uncertainty about the upper bound, since the posterior coefficient of variation CV is much lower than the prior one.

Several additional analyses can be made using posterior statistics, as shown in Fernandes et al. (2010). Assessment of the quantile curve uses the previously discussed concepts of the Monte Carlo method, particularly Eq. (11.23). The quantile curve together with its respective 95% HPD credibility intervals are shown in Fig. 11.16.

Concerning the method proposed by Fernandes et al. (2010), some considerations are important to understand the context of the described application. The authors analyzed a wide range of models and approaches, including the following:

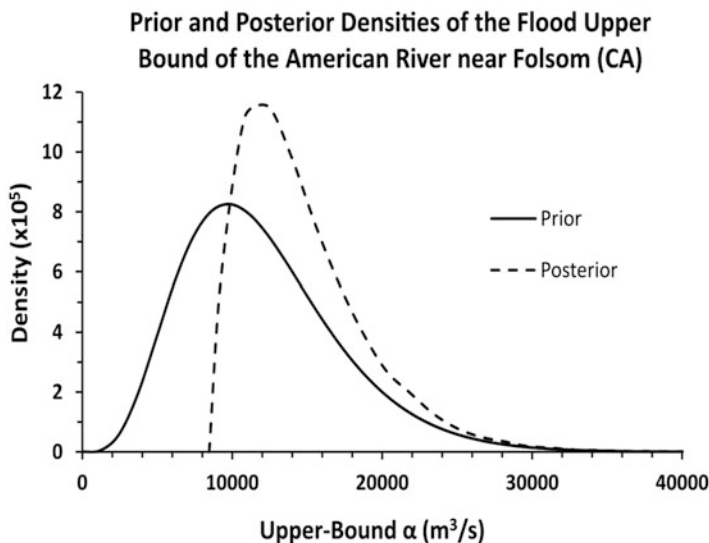


Fig. 11.15 Prior and posterior densities of the upper bound

Table 11.1 Prior and posterior summaries for the upper-bound α

Model	Mean	CV	95 % HPD
Posterior using all data	28,104	0.12	(24,070; 34,980)
Posterior using only systematic data	14,628	0.29	(8609; 22,960)
Prior information	12,018	0.40	

- The likelihood function of the LN4 distribution and of two other upper bounded distributions;
- The likelihood function with and without nonsystematic data;
- Different prior distributions, including non-informative distribution; and
- Parameter estimation under the Bayesian and the classical approaches.

To summarize, results show that the incorporation of nonsystematic data significantly improves estimation of extreme quantiles. They also show that the elicitation of an informative prior distribution is essential when analyzing very rare floods. Furthermore, it was shown that the Bayesian framework is able to combine the objective and subjective aspects of flood frequency analysis.

11.6 Further Reading and Software

This is an entry-level chapter on Bayesian methods, with a particular focus on hydrological applications. For a deeper grasp of Bayesian statistics, interested readers are referred to Bernardo and Smith (1994), Migon and Gamerman (1999),

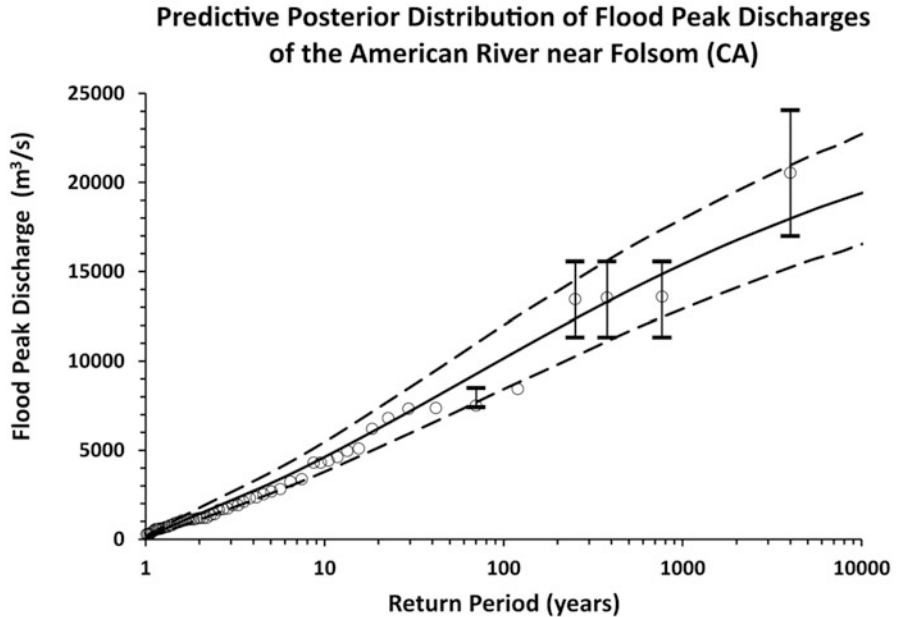


Fig. 11.16 Predictive posterior distribution of flood quantiles (*continuous line*); 95 % credibility intervals (*dashed line*). Circles represent systematic data and circles within bars represent nonsystematic data

Gelman et al. (2004), Paulino et al. (2003) and Robert (2007). In Chap. 11 of his textbook, entitled “A defense of the Bayesian choice,” Robert (2007) makes a point-by-point justification of the Bayesian approach including rebuttals of the most common criticisms of the Bayesian paradigm, which makes for a particularly interesting reading. Renard et al. (2013) provide an up-to-date overview of Bayesian methods for frequency analysis of hydrological extremes with an emphasis on nonstationarity analysis.

There are many software packages available for MCMC. In R, there are the `mcmc` (<http://www.stat.umn.edu/geyer/mcmc/>, accessed: 26th March 2016), `MCMCpack` (Martin et al. 2011) and `LaplacesDemon` (currently available on <https://github.com/ecbrown/LaplacesDemon>, accessed: 26th March 2016) packages. Alternatively there are the `WinBUGS` (Lunn et al. 2000) and `JAGS` (<http://mcmc-jags.sourceforge.net/>, accessed: 26th March 2016) software packages.

Exercises

1. Solve Example 11.1 considering that 1 year after the factory operation started, it was verified that all the floods lasted for more than 5 days, that is, $\varepsilon = 1$. Calculate the posterior distribution.

2. Solve Example 11.2 considering, (a) $s_p^2 = 0.1$ and (b) $s_p^2 = 0.001$. Plot the posterior distribution and comment on the uncertainty of the parameter in each case.
3. The daily number of ships docking at a harbor is Poisson-distributed with mean θ , whose prior distribution is Exponential with mean 1. Knowing that in a 5-day stretch the number of arrivals was 3, 5, 4, 3, and 4: (a) determine the posterior distribution of θ ; and, (b), obtain the 90 and 05 % credibility intervals for θ (Adapted from Paulino et al. 2003).
4. Show that, for a Normal distribution with known σ , if $\mu_{\text{prior}} \sim N(\mu_\mu, \sigma_\mu)$ then

$$\mu_{\text{posterior}} \sim N\left(\frac{\mu_\mu(\sigma^2/n) + \bar{x}\sigma_\mu^2}{\sigma^2/n + \sigma_\mu^2}, \sqrt{\frac{\sigma_\mu^2(\sigma^2/n)}{\sigma_\mu^2 + (\sigma^2/n)}}\right).$$
5. Two meteorologists, A and B, wish to determine the annual rainfall depth (θ in mm) over an ungauged region. Meteorologist A has the prior belief $\theta \sim N(1850, 30^2)$ while meteorologist B believes that $\theta \sim N(1850, 70^2)$. A rain gauge is installed in a region. After 1 year, the rain gauge registered the cumulative rainfall of $x = 1910$. Find the meteorologists' posterior distributions, considering that the standard error of the annual cumulative rainfall depth is 40 mm and that annual rainfalls are normally distributed, i.e., $X|\theta \sim N(\theta, 40^2)$.
6. Consider a normally distributed variate with mean θ and standard deviation 2. A normal prior for θ , with variance 1, was elicited. What is the minimum size of the sample in order for the posterior standard deviation to be 0.1?
7. Consider that $X \sim \text{Ge}(\theta)$. Obtain the Jeffreys prior for θ .
8. Consider the elicitation of the prior distribution for the upper bound as described in Sect. 11.5.4. Suppose further that the upper bound $\text{CV} = 0.3$ and that the local PMF is 25,655 m³/s. Elicit the prior distribution of the upper bound based on the gamma distribution and in the following settings: (a) there is very strong evidence that the PMF will be exceeded in the future; (b) there is a very strong evidence that the PMF will not be exceeded in the future; and (c), there is no evidence regarding the probability of the PMF.
Hint: for the Gamma distribution with parameters ρ_α and β_α , the combination of equations of the method of moments results in $\rho_\alpha = \text{CV}^{-2}$. Parameter β_α can be estimated by attributing a non-exceedance probability p to the current PMF estimate, that is, $P(\alpha \leq \text{PMF} | \rho_\alpha, \beta_\alpha) = p$.
9. Using the WinBUGS algorithm shown in Sect. 11.5.5, and the prior distributions elicited in Exercise 8, find the posterior distributions of the upper bound, considering that the remaining parameters of the LN4 distribution have non-informative priors.
10. Solve Exercises 8 and 9 for $\text{CV} = 0.7$.
11. Consider the sample of annual maximum flows of the Lehigh River at Stoddartsville, listed in Table 7.1. Fit the GEV distribution to the data, with the geophysical prior for the shape parameter, using the following WinBUGS code:

```

# GEV distribution

model {

  for (i in 1:NEX) {
    index[i] <- 1-equals(step(x[i]-mi-sigma/(p-0.5)), step((p-0.5)))
    index2[i] <- equals(index[i], 1)+1
    L[i,1] <- 0
    Z[i] <- index[i]*((p-0.5)/sigma)*(x[i]-mi)
    L[i,2] <- index[i]*(1/sigma)*pow(1-Z[i], 1/(p-0.5)-1)*exp(-pow(1-Z
    [i], 1/(p-0.5)))
  }
  for (i in 1:NEX) {
    dummy[i] <- 0
    dummy[i] ~ dgeneric(phi[i])
    # phi(i) = log(likelihood)
    phi[i] <- log(L[i, index2[i]])
  }

  # priors
  sigma ~ dgamma(1.0, 1.0E-8)
  mi ~ dnorm(1.0, 1.0E-6)
  p ~ dbeta(6.0, 9.0)
}

```

12. Solve Exercise 11 for the Gumbel distribution, using non-informative priors for both parameters.

References

- Albert J (2009) Bayesian computation with R, 2nd edn. Springer, New York
- Anderson H (1986) Metropolis, Monte Carlo, and The MANIAC. Los Alamos Science. U.S. Government Printing Office, pp 96–108
- Ang A, Tang W (2007) Probability concepts in engineering, 2nd edn. Wiley, Hoboken
- Bernardo J, Smith A (1994) Bayesian theory. Wiley, Chichester, UK
- Brooks S (2003) Bayesian computation: a statistical revolution. *Philos Trans Roy Soc A Math Phys Eng Sci* 361(1813):2681–2697
- Coles S, Powell E (1996) Bayesian methods in extreme value modelling: a review and new developments. *Int Stat Rev* 64(1):119
- Dawdy D, Lettenmaier D (1987) Initiative for risk-based flood design. *J Hydraul Eng* 113 (8):1041–1051
- Dubler J, Grigg N (1996) Dam safety policy for spillway design floods. *J Prof Issues Eng Educ Pract* 122(4):163–169
- Ehlers R (2016) Inferência Bayesiana (Notas de Aula) Ricardo Ehlers. <http://www.icmc.usp.br/~ehlers/bayes/>. Accessed 1st Apr 2016

- FEMA—Federal Emergency Management Agency (2004) Federal guidelines for dam safety: selecting and accommodating inflow design floods for Dams. The Interagency Committee of the U.S. Department of Homeland Security, Washington
- Fernandes W (2009) Método para a estimação de quantis de enchentes extremas com o emprego conjunto de análise bayesiana, de informações não sistemáticas e de distribuições limitadas superiormente. PhD Thesis. Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
- Fernandes W, Naghettini M, Loschi R (2010) A Bayesian approach for estimating extreme flood probabilities with upper-bounded distribution functions. *Stoch Environ Res Risk Assess* 24 (8):1127–1143
- Francés F (2001) Incorporating Non-Systematic Information to Flood Frequency Analysis Using the Maximum Likelihood Estimation Method. In: Glade T, Albini P, Francés F (eds) *The use of historical data in natural hazard assessments*, 1st edn. Springer, Dordrecht, pp 89–99
- Francis F, Salas J, Boes D (1994) Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. *Water Resour Res* 30(6):1653–1664
- Gamerman D, Lopes H (2006) *Markov chain Monte Carlo*, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Gelman A, Carlin J, Stern H, Rubin D (2004) *Bayesian data analysis*, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Gilks W, Richardson S, Spiegelhalter D (1996) *Markov chain Monte Carlo in practice*. Chapman & Hall, London
- Hastings W (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109
- Hitchcock D (2003) A history of the Metropolis-Hastings algorithm. *Am Stat* 57(4):254–257
- Hosking J, Wallis J (1997) *Regional frequency analysis*. Cambridge University Press, Cambridge
- ICOLD—International Commission of Large Dams (1987) *Dam safety guidelines*. Bulletin 59. ICOLD (International Congress of Large Dams), Paris
- Liu J (2001) *Monte Carlo strategies in scientific computing*. Springer, New York
- Lunn D, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist Comput* 10:325–337
- Martin A, Quinn K, Park J (2011) MCMCpack: Markov Chain Monte Carlo in R. *J Stat Software* 42(9)
- Martins E, Stedinger J (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour Res* 36(3):737–744
- McGrayne S (2011) *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. Yale University Press, New Haven, CT
- Metropolis N (1987) *The beginning of the Monte Carlo Method*. Los Alamos Science. U.S. Government Printing Office, Special Issue, 1986-676-104/40022
- Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092
- Migon H, Gamerman D (1999) *Statistical inference: an integrated approach*. Arnold, London
- Naulet R (2002) *Utilisation de l'information des crues historiques pour une meilleure prédétermination du risque d'inondation. Application au bassin de l'Ardèche à Vallon Pont-d'Arc et St-Martin d'Ardèche*. PhD Thesis. INRS-ETE
- Paulino C, Turkman M, Murteira B (2003) *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa
- Renard B, Sun X, Lang M (2013) Bayesian Methods for Non-stationary Extreme Value Analysis. In: AghaKouchak A, Easterling D, Hsu K, Schubert S, Sorooshian S (eds) *Extremes in a changing climate: detection, analysis and uncertainty*, 1st edn. Springer, Dordrecht, pp 39–95
- Robert C (2007) *The Bayesian choice*, 2nd edn. Springer, New York
- Robert C, Casella G (2004) *Monte Carlo statistical methods*. Springer, New York

- Silva A, Portela M, Naghettini M, Fernandes W (2015) A Bayesian peaks-over-threshold analysis of floods in the Itajaí-açu River under stationarity and nonstationarity. *Stoch Environ Res Risk Assess* 1:1–20. doi:[10.1007/S00477-015-1184-4](https://doi.org/10.1007/S00477-015-1184-4)
- Stedinger J, Cohn T (1986) Flood frequency analysis with historical and paleoflood information. *Water Resour Res* 22(5):785–793
- Takara K, Tosa K (1999) Storm and flood frequency analysis using PMP/PMF estimates. In: *International Symposium on Floods and Droughts*. Nanjing, pp 7–17
- USBR—United States Bureau of Reclamation (2002) Flood hazard analysis, Folsom dam. Central Valley Project. USBR, Denver
- USBR—United States Bureau of Reclamation (2004) Hydrologic hazard curve estimating procedures. Research Report DSO-04-08. USBR, Denver
- USNRC—United States Nuclear Regulatory Commission (1977) Design basis floods for nuclear power plants. Regulatory Guide 1.59. USNRC, Washington
- Viglione A, Merz R, Salinas J, Blöschl G (2013) Flood frequency hydrology: 3. A Bayesian analysis. *Water Resour Res* 49(2):675–692
- WMO—World Meteorological Organization (1986) Manual for estimation of probable maximum precipitation. Operational Hydrologic Report No. 1, WMO No. 332, 2nd edn. WMO, Geneva