Chapter 3
Bayes Factor for Evaluative Purposes

3.1 Introduction

Consider a case where material of known source (control material) and evidential material of unknown source (recovered or questioned material) are collected and analyzed. Interpretation of scientific evidence then amounts to assessing the probative value of the observations made during comparative examinations. The evidence is evaluated in terms of its effect on the odds in favor of a proposition H_1 put forward by the prosecution, compared to an alternative proposition H_2 advanced by the defense.

During comparative examinations, observations and measurements are made, leading to either discrete or continuous data. Forensic laboratories may also have equipment and methodologies that can lead to output in the form of multivariate data. Thus, scientific evidence is often described by more than one variable. For example, glass fragments from a crime scene can be compared with fragments collected on the clothing of a person of interest on the basis of several chemical components, as well as physical characteristics. It should be noted, however, that the assessment of a Bayes factor for multivariate data may be challenging. For example, data may not present enough regularity so that standard parametric distributions cannot be used. Data may also present a complex dependence structure with several levels of variation. In addition, a feature-based approach might not be always feasible, and it may be necessary to derive a Bayes factor on the basis of scores.

This chapter is structured as follows. Sections 3.2 and 3.3 address the problem of evaluation of evidence for various types of discrete and continuous data, respectively. Section 3.4 presents an extension to continuous multivariate data.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-09839-0_3. The files can be accessed individually by clicking the DOI link in the accompanying figure caption or by scanning this link with the SN More Media App.

S. Bozza et al., *Bayes Factors for Forensic Decision Analyses with R*,
Springer Texts in Statistics, https://doi.org/10.1007/978-3-031-09839-0_3

3.2 Evidence Evaluation for Discrete Data

This section deals with measurement results in the form of counts, using the binomial model (Sect. 3.2.1), the multinomial model (Sect. 3.2.2), and the Poisson model (Sect. 3.2.3).

3.2.1 Binomial Model

In many practical applications, data derive from realizations of experiments that may take one of two mutually exclusive outcomes. Examples include general features (so-called class characteristics) observed on questioned and known items or materials (e.g., fired bullets, fibers) when the question of interest is whether the compared materials come from the same source.

Consider a hypothetical case involving a questioned document for which results of analyses of black toner are available. On the questioned document, black bi-component toner is present. It is of the same type as that used by a given printing machine (known source). A question that may be of interest in such a case is how this analytical information should affect one's belief in the proposition according to which the questioned document has been printed using the device of interest (Biedermann et al., 2009, 2011a). The competing propositions can thus be defined as follows:

H_1 : The questioned document has been printed with the device of interest.
H_2 : The questioned document has been printed with an unknown device.

Let T denote the observed toner type, either single component (T_S) or bi-component (T_B). Suppose that a database of the toner type (magnetism) of samples of black toner from N machines is available, n of which use a bi-component toner. Denote by θ the proportion of the population of printing devices equipped with bi-component toner. Available counts can be treated as realizations of Bernoulli trials (Sect. 2.2.1) with constant probability of success θ, $\Pr(T_B \mid \theta) = \theta$. Suppose a conjugate beta prior distribution $\text{Be}(\alpha, \beta)$ is used to model uncertainty about θ, where α and β can be elicited using the available background knowledge as in (1.42) and (1.43).

Denote by E_y the observations made on recovered material and by E_x the observations made on control material (i.e., documents printed with the device of interest). If the questioned document originates from the device of interest, the probability of the evidence becomes

$$\begin{aligned}\Pr(E_y = T_B, E_x = T_B \mid H_1) &= \int_\Theta \Pr(T_B \mid \theta) \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta / \text{B}(\alpha, \beta) \\ &= \int_\Theta \theta \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta / \text{B}(\alpha, \beta).\end{aligned}$$

If the questioned document originates from an unknown device (i.e., two distinct devices have been used), the probability of the evidence becomes

$$\Pr(E_y = T_B, E_x = T_B \mid H_2) = \int_{\Theta} \theta^2 \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta / \mathrm{B}(\alpha, \beta).$$

The Bayes factor can be computed as

$$\begin{aligned} \mathrm{BF} &= \frac{\int_{\Theta} \theta \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta}{\int_{\Theta} \theta^2 \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta} \\ &= \frac{\mathrm{B}(\alpha+1, \beta)}{\mathrm{B}(\alpha+2, \beta)} \int_{\Theta} \frac{\theta^{\alpha}(1-\theta)^{\beta-1}}{\theta^{\alpha+1}(1-\theta)^{\beta-1}} \frac{\mathrm{B}(\alpha+2, \beta)}{\mathrm{B}(\alpha+1, \beta)} \\ &= \frac{\alpha+\beta+1}{\alpha+1}. \end{aligned} \tag{3.1}$$

Example 3.1 (Questioned Documents) Consider the case of a printed document of unknown origin. Analyses reveal that the toner present on the printed document is of type "bi-component." The printing device that is thought to have been used to print the questioned document is equipped with a bi-component toner. In an available database with a total of $N = 100$ samples of black toner, $n = 23$ are bi-component (see Table 3.1). Using this information, the parameters of the beta prior distribution about θ can be elicited as follows:

```
> n=23
> N=100
> p=n/N
> a=p*(N-1)
> b=(1-p)*(N-1)
```

This leads to a Be(23, 76).

The Bayes factor in (3.1) can be computed straightforwardly as follows:

```
> BF=(a+b+1)/(a+1)
> BF

[1] 4.206984
```

The Bayes factor provides weak support for the proposition H_1 according to which the questioned document has been printed with the printing device of interest rather than with an unknown printing device (H_2).

It is worth noting that there is an alternative development described in the forensic statistics literature that considers background information derived from a population database as part of the evidence, (e.g., Ommen et al., 2016; Dawid,

Table 3.1 Results obtained following the analysis of, respectively, the component type (magnetism) and the resin type of 100 samples of black toner (Biedermann et al., 2011a)

Resin group	Single component	Bi-component
1. Styrene-co-acrylate	69	14
2. Epoxy A	8	3
3. Epoxy B	0	2
4. Epoxy C	0	1
5. Epoxy D	0	1
6. Polystyrene	0	1
7. Other	0	1

2017). According to this line of reasoning, if proposition H_1 is true (numerator), there are $(n + 1)$ counts of bi-component toners. That is, the questioned item and the known item are assumed to come from the same source, hence adding one count to the database. Conversely, if proposition H_2 is true (denominator), there are $(n+2)$ counts of bi-component toner. Here, it is assumed that the questioned item and the known item come from different sources, hence adding two counts to the database. The Bayes factor can then be obtained as

$$\begin{aligned} \text{BF} &= \frac{\int_\Theta \theta^{n+1}(1-\theta)^{N-n}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta}{\int_\Theta \theta^{n+2}(1-\theta)^{N-n}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta} \\ &= \frac{\alpha+\beta+N+1}{\alpha+n+1}. \end{aligned} \tag{3.2}$$

One can immediately verify that this corresponds to the BF in (3.1) with parameter α replaced by $\alpha + n$, and parameter β replaced by $\beta + N - n$. However, it may be questioned whether the available database should be considered as evidence, rather than as conditioning information, because the database contains only general data unrelated to the case under investigation (Aitken et al., 2021).

3.2.2 Multinomial Model

The analyses described in Sect. 3.2.1 can be extended to situations where experiments can lead to more than two mutually exclusive outcomes.

Consider again the case involving printed documents, introduced in Sect. 3.2.1. Laboratories often analyze resins of toner on printed documents by means of Fourier Infrared Spectroscopy (FTIR). The results can be classified into one of several (k) categories (Table 3.1). Suppose that the resin type (R) recovered on the questioned document belongs to category j, which is also found in the toner used by a given printing machine. The question of interest is similar to the one considered in Sect. 3.2.1, that is, how the available analytical information should affect one's belief in the proposition according to which a questioned document has been printed using a given device, called the potential source, rather than by some unknown printing device.

Denote by θ_j the proportion of the population that is of type (category) R_j, $j = 1, \ldots, k$, $\Pr(R_j \mid \theta_j) = \theta_j$. Assume that observations of distinct categories can be treated as independent: available counts $n_1, \ldots, n_k$ can be treated as realizations from a multinomial distribution $\text{Mult}(n, \theta_1, \ldots, \theta_k)$

$$f(n_1, \ldots, n_k \mid \theta_1, \ldots, \theta_k) = \frac{N!}{n_1! \cdot \cdots \cdot n_k!} \theta_1^{n_1} \cdot \cdots \cdot \theta_k^{n_k}.$$

A conjugate Dirichlet prior probability distribution $\text{Dir}(\alpha_1, \ldots, \alpha_k)$ is considered for modeling uncertainty about the population proportions $\theta_1, \ldots, \theta_k$:

$$f(\theta_1, \ldots, \theta_k \mid \alpha_1, \ldots, \alpha_k) = \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} / \mathrm{B}(\alpha),$$

with $\mathrm{B}(\alpha) = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma(\alpha)}$ and $\alpha = \sum_{i=1}^k \alpha_i$.

Denote by E_y the observations made on the recovered material and by E_x the observations made on the control material (i.e., documents printed with the device of interest). If the questioned document originates from the device of interest, the probability of the findings $E = (E_y, E_x)$ becomes

$$\begin{aligned}\Pr(E_y = R_j, E_x = R_j \mid H_1) &= \int_\Theta \Pr(R_j \mid \theta_j) \cdot \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} / \mathrm{B}(\alpha) \\ &= \int_\Theta \theta_j \cdot \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} / \mathrm{B}(\alpha).\end{aligned}$$

If the questioned documents originate from an unknown device (i.e., two distinct devices have been used), the probability of the findings E becomes

$$\Pr(E_y = R_j, E_x = R_j \mid H_2) = \int_\Theta \theta_j^2 \cdot \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} d\boldsymbol{\theta} / \mathrm{B}(\alpha).$$

The Bayes factor can be computed as

$$\begin{aligned}\mathrm{BF} &= \frac{\int \theta_j \cdot \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} d\boldsymbol{\theta}}{\int \theta_j^2 \cdot \theta_1^{\alpha_1 - 1} \cdot \cdots \cdot \theta_j^{\alpha_j - 1} \cdot \cdots \cdot \theta_k^{\alpha_k - 1} d\boldsymbol{\theta}} \\ &= \frac{\alpha + 1}{\alpha_j + 1}.\end{aligned} \tag{3.3}$$

Example 3.2 (Questioned Documents—Continued) Recall Example 3.1, involving questioned documents on which black toner is present. Suppose now that laboratory analyses focus on the toner's resin component. Suppose that the parameters of the Dirichlet prior probability distribution are elicited as

```
> a=c(15,4,3,2,2,2,2)
```

Suppose that the rather common resin group *Epoxy-A* (category $j = 2$ in Table 3.1) is observed on both the questioned and known documents. The Bayes factor in (3.3) can be computed straightforwardly as

```
> j=2
> BF=(sum(a)+1)/(a[j]+1)
> BF
```

```
[1] 6.2
```

The Bayes factor provides, again, weak support for the proposition H_1 according to which the questioned document has been printed with the printing device of interest, rather than with an unknown printing device (H_2).

Suppose that a database of the resin type of samples of black toner from N machines is available, n_1 ($n_2, \ldots$) of which belong to category 1 (2, ...), as in Table 3.1. These data can be used to elicit the Dirichlet prior probability distribution. Following the methodology proposed by Zapata-Vazquez et al. (2014), the hyperparameters $\alpha_1, \ldots, \alpha_k$ can be assessed by starting from expert judgments (e.g., a vector of quantiles) about proportions of items belonging to each category. Tools for eliciting prior probability distributions from experts' opinions are also available in the `R` package `SHELF`. An example will be presented in Sect. 4.2.2.

3.2.3 Poisson Model

Some forensic science applications focus on the number of occurrences of particular events or observations that take place at given intervals of time or space. Practical examples are the number of gunshot residue particles (GSR) collected on the surface of the hands of individuals suspected to be involved in the discharge of a firearm (Cardinetti et al., 2006), or the number of corresponding matching striations in the comparative examination of marks left by firearms on fired bullets (Bunch, 2000).

Consider the following hypothetical case. A fired bullet is found at a crime scene, and a person of interest is apprehended, carrying a gun. The following propositions are of interest:

H_1 : The bullet found at the crime scene was fired with the seized gun.
H_2 : The bullet found at the crime scene was fired with an unknown gun.

The recovered bullet and bullets fired with the seized gun are compared. *Consecutive matching striations* (CMS) is a simple concept to quantify the extent of agreement between marks. The number of observed consecutively matching striations can be interpreted as a *score*. Let $\Delta(x, y)$ be the maximum CMS count for a given comparison. For the evaluation of a CMS count, data on comparisons made between pairs of bullets test-fired with the seized gun and between pairs of bullets test-fired with different guns are needed. The (score-based) Bayes factor therefore is

$$\text{sBF} = \frac{g(\Delta(x, y) \mid H_1)}{g(\Delta(x, y) \mid H_2)}.$$

A statistical model commonly used in the forensic science literature for the type of data encountered in the example here assumes that counts follow a Poisson distribution $\text{Pn}(\lambda)$

$$g(\Delta(x, y) \mid \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{\Delta(x,y)}}{\Delta(x, y)!}, \qquad \Delta(x, y) = 0, 1, \ldots ; \ \lambda_i \geq 0,$$

where parameter λ_i, $i = 1, 2$, represents the weighted average maximum CMS count.

Suppose that two datasets are compiled. The first relates to pairs of bullets fired with the seized gun, and the second to pairs of bullets fired with different guns. Such data can be used to inform the probability distribution $g(\cdot)$ at the score value $\Delta(x, y)$ as discussed in Sect. 1.5.2 and to compute the Bayes factor as

$$\text{sBF} = \frac{\hat{g}(\Delta(x, y) \mid x, H_1)}{\hat{g}(\Delta(x, y) \mid H_2)}.$$

Bunch (2000) describes a likelihood ratio procedure for inference about competing propositions. This account is based on a frequentist perspective because it uses the maximum likelihood estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ for parameters λ_1 and λ_2, calculated under the assumption that either proposition H_1 or proposition H_2 is true. Using these two estimates in the component Poisson likelihoods leads to the following likelihood ratio:

$$\text{LR} = \frac{e^{-\hat{\lambda}_1} \hat{\lambda}_1^{\Delta(x,y)}}{e^{-\hat{\lambda}_2} \hat{\lambda}_2^{\Delta(x,y)}}.$$

In Bayesian statistics, the most common prior distribution for λ_i is the gamma distribution $\text{Ga}(\alpha_i, \beta_i)$ with shape parameter α and rate parameter β (e.g. Bernardo and Smith, 2000):

$$f(\lambda_i \mid \alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i}, \qquad \lambda_i > 0\,;\ \alpha_i, \beta_i > 0.$$

Since the Poisson and gamma distributions are conjugate (Sect. 1.10), the posterior distribution of λ is still in the family of gamma distributions, with parameters α and β updated according to well-known updating rules (see, e.g., Lee, 2012). When we have a realization of a random sample from a Poisson distribution, Pn(λ), say $(z_1, \ldots, z_n)$, we end up with a Ga(α', β'), where $\alpha' = \alpha + \sum_{i=1}^{n} z_i$ and $\beta' = \beta + n$. Note that in this case there is only one observation, $\Delta(x, y)$; therefore, $\alpha' = \alpha + \Delta(x, y)$ and $\beta' = \beta + 1$. See also Biedermann et al. (2011b) for further illustrations of the Poisson–gamma model in forensic science applications.

The marginal distribution in the numerator and denominator of the Bayes factor is known in closed form here. It is a Poisson–gamma distribution:

$$\begin{aligned} g(\Delta(x, y)|\alpha_i, \beta_i) &= \int_{\lambda_i} g(\Delta(x, y)|\lambda_i) f(\lambda_i|\alpha_i, \beta_i) d\lambda_i \\ &= \frac{1}{\Delta(x, y)!} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \frac{\Gamma(\alpha_i + \Delta(x, y))}{(\beta_i + 1)^{\alpha_i + \Delta(x, y)}}. \end{aligned} \tag{3.4}$$

The score-based Bayes factor then becomes

$$\text{sBF} = \frac{\beta_1^{\alpha_1} \Gamma(\alpha_2) \Gamma(\alpha_1 + \Delta(x, y)) (\beta_2 + 1)^{\alpha_2 + \Delta(x, y)}}{\beta_2^{\alpha_2} \Gamma(\alpha_1) \Gamma(\alpha_2 + \Delta(x, y)) (\beta_1 + 1)^{\alpha_1 + \Delta(x, y)}}. \tag{3.5}$$

Another example of the use of the Poisson distribution for data in the form of independent counts can be found in Aitken and Gold (2013). These authors considered the number of occurrences of selected characteristics of speech recorded in a succession of time periods. In this application, a feature-based Bayes factor is used to assess findings with respect to the proposition according to which recorded and control speeches originate from the same source versus the alternative proposition that they originate from different sources.

Example 3.3 (Firearm Examination) Consider a case involving a questioned bullet. During comparison with a reference bullet, the examiner counts four CMS, i.e., $\Delta(x, y) = 4$. Suppose that the assumptions made in Bunch (2000) are suitable for the case here so that for bullets fired from the same gun (proposition H_1 holds), the weighted average maximum CMS is taken to be equal to 3.91. For bullets fired from different guns (proposition H_2 holds), the weighted average maximum CMS count is taken to be equal to 1.32. These values are used in the Poisson likelihoods under H_1 and H_2, and the likelihood ratio can easily obtained as

```
> s=4
> lambda1=3.91
```

(continued)

Example 3.3 (continued)

```
> lambda2=1.32
> LR=dpois(s,lambda1)/dpois(s,lambda2)
> LR
```

```
[1] 5.775487
```

The evidence provides weak support in favor of the proposition according to which the recovered bullet passed through the barrel of the seized gun, rather than through the barrel of an unknown gun.

Consider now the Bayesian perspective. Suppose that the available knowledge allows one to set the hyperparameters of the gamma distribution equal to $\{\alpha_1 = 125, \beta_1 = 32\}$ for the numerator and to $\{\alpha_2 = 7, \beta_2 = 5\}$ for the denominator. This amounts to using a gamma prior distribution for λ_1 with mean equal to 3.91 and standard deviation equal to 0.35 and a gamma prior distribution for λ_2 with mean equal to 1.4 and standard deviation equal to 0.53. The two prior distributions are shown in Fig. 3.1.

```
> an=125
> bn=32
> ad=7
> bd=5
> plot(function(x) dgamma(x,an,bn),0,8,
+ xlab=expression(paste(lambda)),ylab='Probability
+ density')
> plot(function(x) dgamma(x,ad,bd),0,8,add=TRUE,
+ lty=2)
> leg=expression(paste('Ga(125,32)'),paste(
+'Ga(7,5)'))
> legend(4.85,1.15,leg,lty=c(1,2))
```

First, we write a short function `poisg` that computes the marginal distribution in (3.4)

```
> poisg=function(a,b,x)
+ {(b^a)/gamma(a)*gamma(a+x)/((b+1)^(a+x))}
```

Next, the Bayes factor can be computed as follows:

```
> BF=poisg(an,bn,s)/poisg(ad,bd,s)
> BF
```

```
[1] 4.248019
```

Note that the introduction of a prior probability distribution reflecting uncertainty about the population parameters λ_1 and λ_2 has slightly lowered the value of the evidence. The result still represents weak evidence in favor of the

(continued)

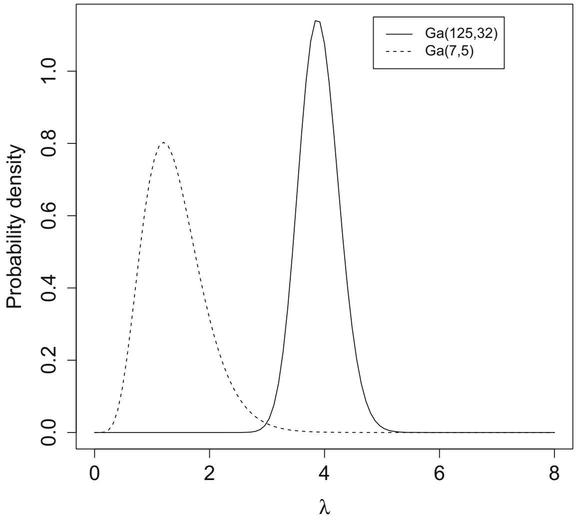

Fig. 3.1 Gamma prior for the Poisson parameter λ under H_1 (solid line) and H_2 (dashed line)

Example 3.3 (continued)
proposition that the recovered bullet was fired with the seized gun, rather than with an unknown gun.

Note that Example 3.3 involves a non-anchored approach at the numerator. The probability distribution of the score value is solely conditioned on the hypothesis of interest, that is $\hat{g}(\Delta(x, y) \mid H_1)$. As mentioned at the beginning of this section, and in Sect. 1.5.2, other anchoring approaches may be considered.

3.2.3.1 Choosing the Parameters of the Gamma Prior

An evaluator who, initially, would like to give the same weight to all possible values of λ may consider to use a non-informative prior distribution, that is

$$f(\lambda_i) = \lambda_i^{-1/2}; \qquad \lambda_i > 0 \text{ and } i = 1, 2.$$

The posterior probability distribution given the observations $(z_1, \ldots, z_n)$ will be of type gamma with shape parameter $\alpha' = \sum_{i=1}^{n} z_i + 1/2$ and rate parameter $\beta' = n$. Note that in the type of case considered here, there is only one observation; therefore, $\alpha' = \Delta(x, y) + 1/2$ and $\beta' = 1$.

However, the choice of a non-informative prior distribution may be questioned. Take, for instance, the case example discussed earlier in this section (Example 3.3). It is difficult to imagine that *no* suitable information is available to express prior uncertainty about the unknown weighted average maximum count CMS, and hence that the same non-informative prior distribution should apply under each proposition.

In Example 3.3, an informative prior distribution has been used. This raises the question of how to translate prior knowledge into a prior distribution. As illustrated in Sect. 1.10, one way to elicit prior parameters is to express prior beliefs in terms of a measure of location and a measure of dispersion and then equate these values with the prior moments of the distribution. In the case of a gamma distribution $\text{Ga}(\alpha, \beta)$, this amounts to equate a value for the mean, m, with the prior mean α/β, and a value for the variance, s^2, with the prior variance α/β^2, that is,

$$m = \frac{\alpha}{\beta} \qquad ; \qquad s^2 = \frac{\alpha}{\beta^2}.$$

Solving for α and β gives

$$\alpha = \frac{m^2}{s^2} \tag{3.6}$$

$$\beta = \frac{m}{s^2}. \tag{3.7}$$

If the shape of the prior distribution resulting from the choice of α and β as in (3.6) and (3.7) does not reflect one's prior beliefs suitably, then one should adjust the numerical values of m and s. However, this may not be enough to ensure that the resulting prior distribution is reasonable. One should also inquire about whether the information that is conveyed by the prior is realistically attainable. Consider a random sample of size n_e, providing the same amount of information as conveyed by the elicited prior. The sample mean should have, at least roughly, the same location and the same dispersion as the prior. The equivalent sample size n_e can then be found by matching the moments of the gamma distribution to the corresponding moments characterizing a sample of size n_e from a Poisson distributed random variable located at λ:

$$\frac{\alpha}{\beta} = \lambda$$

$$\frac{\alpha}{\beta^2} = \frac{\lambda}{n_e}.$$

If the mean λ is set equal to the prior mean α/β, the equivalent sample size n_e is equal to β.

Example 3.4 (Elicitation of a Gamma Prior) In Example 3.3, a Ga(125, 32) was used for λ_1 (the weighted average maximum CMS count under proposition H_1), and a Ga(7, 5) for λ_2 (the weighted average maximum CMS count under proposition H_2). For the prior means of λ_1 and λ_2, the values 3.91 and 1.4 were used following Bunch (2000). For the dispersion of the two distributions, the values 0.35 and 0.53 have been assigned to the standard deviation under propositions H_1 and H_2, respectively. Parameters ($\alpha_1 = 125, \beta_1 = 32$) and ($\alpha_2 = 7, \beta_2 = 5$) have then been obtained as in (3.6) and (3.7). This amounts to an equivalent sample size equal to 32 for the prior density of λ_1, and 5 for λ_2.

3.2.3.2 Sensitivity to Prior Probabilities of Competing Propositions

It is important to emphasize that the analyses presented here make no direct probabilistic statement about the truth of the propositions put forward by opposing parties at trial. A Bayes factor of approximately 4.25, as obtained in Example 3.3, only means that the evidence is approximately 4 times more probable if proposition H_1 is true than if the alternative proposition H_2 is true. As noted earlier, this does not mean that proposition H_1 is more probable than H_2. This depends on the prior probabilities of the competing propositions, which can vary considerably among recipients of expert information, and which are beyond the area of competence of scientists.

However, it may be of interest to show the impact of different prior probability assignments on the posterior probability of the competing propositions. To do so, recall that the posterior odds are given by the product of the prior odds and the Bayes factor

$$\frac{\Pr(H_1 \mid \cdot)}{\Pr(H_2 \mid \cdot)} = \text{BF} \times \frac{\Pr(H_1)}{\Pr(H_2)}.$$

Using this expression, one can then investigate how the posterior probability of proposition H_1, i.e., α_1, varies for values of π_1, i.e., $\Pr(H_1)$, ranging from 0.01 until 0.99, and for a Bayes factor equal to 4.25, as in Example 3.3.

```
> pi1=seq(0.01,0.99,0.01)
> prior_odds=pi1/(1-pi1)
> BF=4.25
> post_odds=prior_odds*BF
> alpha1=post_odds/(1+post_odds)
```

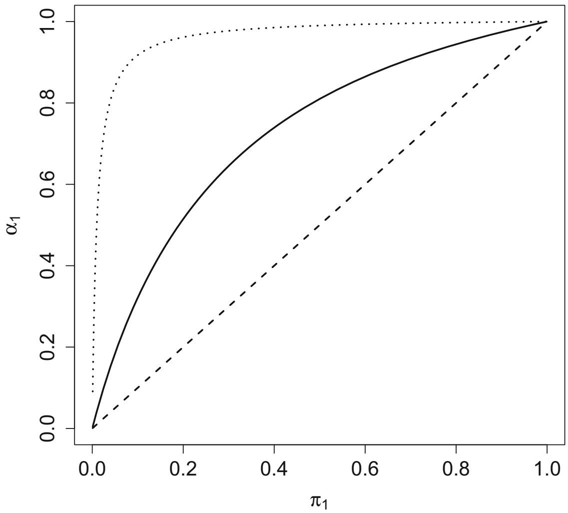

Fig. 3.2 Posterior probability α_1 of proposition H_1 for values of prior probabilities π_1 ranging from 0.01 to 0.99, and a Bayes factor equal to 4.25 (solid line), 1 (dashed line), and 100 (dotted line)

The solid line in Fig. 3.2 shows the value of α_1, the posterior probability of the proposition H_1, as a function of the prior probability, π_1, for BF $= 4.25$. The plot also shows results for BF $= 1$ (dashed line) and for BF $= 100$ (dotted line).

```
> plot(pi1,alpha1,type='l',xlab=expression(pi[1]),
+ ylab=expression(alpha[1]))
> BF=1
> post_odds=prior_odds*BF
> alpha1=post_odds/(1+post_odds)
> lines(pi1,alpha1,lty=2)
> BF=100
> post_odds=prior_odds*BF
> alpha1=post_odds/(1+post_odds)
> lines(pi1,alpha1,lty=3)
```

More generally, it can be observed that the higher the value of the Bayes factor, the smaller the impact of the prior probabilities on posterior probabilities.

3.3 Evidence Evaluation for Continuous Data

The previous section considered the evaluation of scientific evidence as given by discrete data. However, for many types of evidence, measurements result in continuous data.

3.3.1 Normal Model with Known Variance

In some applications, the distribution of measurements exhibits enough regularity to be captured by standard parametric models, such as the Normal distribution. One example, introduced earlier in Sect. 1.5.1, is the analysis of magnetism of black toner on printed documents. Due to the wide distribution and availability of printing machines, forensic document examiners are commonly requested to examine documents produced by electrophotographic printing processes that use dry toner. A question that forensic scientists may be asked to help with is whether or not two or more documents were printed with the same laser printer. This task involves the comparison of analytical features of a questioned document with those of control documents. One such analytical feature is the magnetic flux of toner. It is thought to be largely influenced by individual settings of the printing device, so that detectable differences may be expected on documents printed at different instances using the same or different machines (Biedermann et al., 2016a).

Suspected page substitution is a commonly encountered problem in forensic document examination. Imagine a case involving a contract consisting of three pages where the allegation is that the second page has been substituted. It may be of interest, thus, to investigate the extent to which available measurements of magnetic flux can be informative in this case.

Consider the following pair of propositions:

H_1 : Page two has been printed by the device used for printing pages one and three (i.e., the three pages have been printed with the same device).
H_2 : Page two has been printed by a different device.

Denote by $\mathbf{y} = (y_1, \ldots, y_n)$ the measurements of magnetic flux obtained for the questioned page. Measurements are assumed to be normally distributed with unknown mean θ and known variance σ^2. The likelihood of the normal random sample $(y_1, \ldots, y_n)$ can therefore be expressed as

$$f(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \theta)^2\right\}. \tag{3.8}$$

It can be shown, (e.g., Bolstad and Curran, 2017), that the likelihood of a normal random sample is proportional to the likelihood of the sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The sample mean is normally distributed with mean θ and variance σ^2/n

$$f(\bar{y} \mid \theta) = (2\pi\sigma^2/n)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2/n}(\bar{y} - \theta)^2\right\}. \tag{3.9}$$

In other words, it is possible to reduce the problem to one where a single normal observation $\bar{y}$ is available.

Next, denote the measurements on uncontested pages by $\{\mathbf{x}_l\} = (x_{lj}, j = 1, \ldots, n$ and $l = 1, 2)$, where the subscript l refers to the page number and j to the number of measurements of magnetic flux obtained for the page l. A normal distribution with mean θ and variance σ^2 is assumed for $\mathbf{x}$, analogously to what has been assumed for $\mathbf{y}$. A conjugate normal prior distribution is chosen for θ, say $\theta \sim \mathrm{N}(\mu, \tau^2)$. The Bayes factor can be computed as in (1.16):

$$\begin{aligned}\mathrm{BF} &= \frac{f(\bar{y} \mid \mathbf{x}_1, \mathbf{x}_2, H_1)}{f(\bar{y} \mid H_2)} \\ &= \frac{\int f(\bar{y} \mid \theta) f(\theta \mid \mathbf{x}_1, \mathbf{x}_2, H_1) \mathrm{d}\theta}{\int f(\bar{y} \mid \theta) f(\theta \mid H_2) \mathrm{d}\theta},\end{aligned} \tag{3.10}$$

where $f(\theta \mid \mathbf{x}_1, \mathbf{x}_2, H_1)$ is the posterior distribution of θ, obtained by updating the prior distribution $\mathrm{N}(\mu, \tau^2)$ using the measurements $\mathbf{x}_1$ and $\mathbf{x}_2$. This is a normal distribution, $(\theta \mid \mathbf{x}_1, \mathbf{x}_2) \sim \mathrm{N}(\mu_x, \tau_x^2)$, with posterior mean μ_x and posterior variance τ_x^2, computed according to the updating rules (2.13) and (2.14). Using the result (1.21), one can easily verify that the density in the numerator is still a normal distribution with mean equal to the posterior mean μ_x and variance equal to the sum of the posterior variance τ_x^2 and the population variance σ^2 divided by the sample size n, i.e., $\tau_x^2 + \sigma^2/n$. In the same way, invoking (1.22), the density in the denominator is still a normal distribution with mean equal to the prior mean μ and variance equal to the sum of the prior variance τ^2 and the population variance σ^2 divided by the sample size n, i.e., $\tau^2 + \sigma^2/n$.

Example 3.5 (Printed Documents) Consider the case described above where a forensic document examiner measures the magnetic flux on two uncontested pages 1 and 3 (Biedermann et al., 2016a). The results are $\mathbf{x}_1 = (16, 15, 15)$ and $\mathbf{x}_2 = (16, 15, 16)$. The measurements for the contested page 2 are $\mathbf{y} = (15, 16, 16)$. Previous experiments allow one to assign the value 0.24 for the population standard deviation σ. Based on the available knowledge regarding the magnetic flux of toner on printed documents, the prior mean μ and the prior variance τ^2 for the unknown quantity of magnetic flux are set equal to 17.5 and 3.92^2, respectively. This means that values of the magnetic flux smaller than 6 and greater than 29 are considered, a priori, to be extremely unlikely.

```
> mu=17.5
> tau2=3.92^2
> sigma2=0.24^2
> x=c(16,15,15,16,15,16)
> y=c(15,16,16)
> nx=length(x)
> ny=length(y)
```

(continued)

Example 3.5 (continued)
The posterior distribution $f(\theta \mid \mathbf{x}_1, \mathbf{x}_2)$ can be obtained by a single application of Bayes theorem with the full set of available measurements $(\mathbf{x}_1, \mathbf{x}_2)$. The posterior parameters μ_x and τ_x^2 can be calculated using the function `post_distr` introduced in Sect. 2.3.1.

```
> mupost=post_distr(sigma2,nx,mean(x),mu,tau2)[1]
> mupost
```

```
[1] 15.50125
```

```
> tau2post=post_distr(sigma2,nx,mean(x),mu,tau2)[2]
> tau2post
```

```
[1] 0.009594006
```

The two marginal densities in the numerator and denominator of the BF in (3.10) can be calculated at the sample mean $\bar{y}$. The exact value of the Bayes factor is given by

```
> BF=dnorm(mean(y),mupost,sqrt(tau2post+sigma2/ny))/
+ dnorm(mean(y),mu,sqrt(tau2+sigma2/ny))
> BF
```

```
[1] 16.03199
```

This value represents moderate support for the proposition of page substitution, compared to the proposition of no page manipulation.

3.3.2 *Normal Model with Both Parameters Unknown*

So far, the variance of the distribution of the observations has been assumed to be known, though in many practical situations the mean and the variance are both unknown, and it is necessary to choose a prior distribution for the parameter vector (θ, σ^2). The Bayes factor can be computed as in (1.16):

$$\begin{aligned} \mathrm{BF} &= \frac{f(\mathbf{y} \mid \mathbf{x}, H_1)}{f(\mathbf{y} \mid H_2)} \\ &= \frac{\int f(\mathbf{y} \mid \theta, \sigma^2) f(\theta, \sigma^2 \mid \mathbf{x}, H_1) \mathrm{d}(\theta, \sigma^2)}{\int f(\mathbf{y} \mid \theta, \sigma^2) f(\theta, \sigma^2 \mid H_2) \mathrm{d}(\theta, \sigma^2)}. \end{aligned} \tag{3.11}$$

Consider the case where a conjugate prior distribution for (θ, σ^2) of the form

$$f(\theta, \sigma^2) = f(\theta \mid \sigma^2) f(\sigma^2) \tag{3.12}$$

is chosen. In this distribution, prior beliefs about the population mean θ are calibrated by the scale of measurements of the observations.[1] The conditional distribution $f(\theta \mid \sigma^2)$ is taken to be normal, centered at μ with variance σ^2/n_0, $(\theta \mid \sigma^2) \sim \mathrm{N}(\mu, \frac{\sigma^2}{n_0})$. The parameter n_0 can be thought of as the prior sample size for the distribution of θ. As pointed out in Sect. 2.3.1, it formalizes the size of the sample from a normal population that provides an equivalent amount of information about θ. The distribution $f(\sigma^2)$ is taken to be an S times inverse chi-squared distribution with k degrees of freedom, $\sigma^2 \sim S \cdot \chi^{-2}(k)$. It can be shown that this is equivalent to an inverse gamma distribution with shape parameter $\alpha = k/2$ and scale parameter $\beta = S/2$, $\sigma^2 \sim \mathrm{IG}(\alpha = k/2, \beta = S/2)$. Alternatively, prior uncertainty about dispersion can be formulated in terms of the precision $\lambda^2 = 1/\sigma^2$. The prior distribution of λ^2 becomes a gamma distribution with shape parameter $\alpha = k/2$ and rate parameter $\beta = S/2$, $\lambda^2 \sim \mathrm{Ga}(\alpha = k/2, \beta = S/2)$. For further discussion, see e.g. Bernardo and Smith (2000), Bolstad and Curran (2017) and Robert (2001).

Consider now the posterior distribution of the unknown parameter vector (θ, λ^2) once a vector of observations $\mathbf{x} = (x_1, \ldots, x_n)$ becomes available. It takes the form of a normal–gamma distribution

$$f(\theta, \lambda^2 \mid \mathbf{x}, H_1) = \mathrm{NG}(\mu_n, n', \alpha_n, \beta_n),$$

with

$$\mu_n = \frac{n\bar{x} + n_0\mu}{n + n_0} \qquad ; \qquad n' = n + n_0$$

$$\alpha_n = \alpha + \frac{n}{2};$$

$$\beta_n = \beta + \frac{1}{2}\left[(n-1)s^2 + \frac{n_0 n(\bar{x} - \mu)^2}{n_0 + n}\right],$$

[1] Note that in (3.12) population parameters are not, a priori, independent. Whenever this condition is felt to be too restrictive (see, e.g., Robert (2001)), it is also possible to choose a prior distribution as the product of independent priors, $f(\theta, \sigma^2) = f(\theta) f(\sigma^2)$. In this case, the derivation of the posterior distribution can be more demanding.

and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

If uncertainty about the two unknown parameters is modeled by means of the conjugate prior distribution in (3.12), the integrations in (3.11) have an analytical solution and the BF can be obtained straightforwardly.

Denote by $\mathbf{y} = (y_1, \ldots, y_{n_y})$ a vector of measurements made on questioned material and consider the sample mean $\bar{y} = \sum_{i=1}^{n_y} y_i$. It can be proved that the marginal density $f(\bar{y} \mid \mathbf{x}, H_1)$ in the numerator is a Student t distribution with $2\alpha + n$ degrees of freedom, centered at μ_n, with spread parameter, denoted s_n, equal to

$$s_n = \frac{n_y(n + n_0)}{n + n_0 + n_y}\left(\alpha + \frac{n}{2}\right)\beta_n^{-1}.$$

This can be denoted as $f_1(\bar{y} \mid \mu_n, s_n, 2\alpha + n)$.

The marginal density $f(y \mid H_2)$ in the denominator is a Student t distribution with k degrees of freedom, centered at μ with spread parameter (precision), denoted s_d, equal to

$$s_d = \frac{n_0 n_y}{n_0 + n_y}\alpha\beta^{-1}$$

(Bernardo and Smith, 2000). This can be denoted as $f_2(\bar{y} \mid \mu, s_d, 2\alpha)$.

The Bayes factor can then be computed as

$$\text{BF} = \frac{f_1(\bar{y} \mid \mu_n, s_n, 2\alpha + n)}{f_2(\bar{y} \mid \mu, s_d, 2\alpha)}. \tag{3.13}$$

Choosing the Parameters of the Normal Prior

The use of a conjugate prior distribution for the mean and the variance of a normal distribution raises the question of how to choose the hyperparameters, as the resulting distribution should suitably reflect available prior knowledge. The prior distribution $f(\theta \mid \sigma^2)$ requires one to choose a value for μ, the measure of location, and a value for n_0. The ratio n_0/n characterizes the relative precision of the prior distribution compared to the precision of the observations. If this ratio is very small, the less informative will be the prior distribution, and the closest will be the posterior distribution to that obtained using a non-informative prior distribution. In fact, when n_0/n approaches zero, the limiting form of the marginal distribution of the population mean θ is $\text{N}(\bar{x}, \sigma^2/n)$, which corresponds to the posterior distribution that would be obtained using a non-informative prior distribution (Robert, 2001). For more specific prior beliefs (i.e., concentrated on a limited range of values), a higher value of n_0 should be chosen.

Regarding the prior distribution of σ^2, consider a number of degrees of freedom $k = 20$ so that the prior mass is distributed rather symmetrically. Suppose also that, based on knowledge available from previous experiments, it is considered that values of σ^2 greater or smaller than 0.05 are equally plausible, so $\Pr(\sigma^2 > 0.05) = 0.5$. The parameter S can be elicited by recalling that $\sigma^2/S \sim \chi^{-2}(k)$ and, analogously, $S \cdot \lambda^2 \sim \chi^2(k)$ so

$$\Pr\left(\sigma^2 > 0.05\right) = \Pr\left(S \cdot \lambda^2 < S \cdot 20\right) = 0.5,$$

where $S \cdot 20$ is the quantile of order 0.5 of a χ^2 distributed random variable with $k = 20$ degrees of freedom.

```
> sigma2=0.05
> k=20
> p=0.5
> q=qchisq(p,k)
> q
[1] 19.33743
> S=q*sigma2
```

Parameter S is then equal to

$$S = 19.3374 \times 0.05 \approx 1.$$

The elicited prior distribution for σ^2 is $\mathrm{IG}(\frac{20}{2}, \frac{1}{2})$ and is shown in Fig. 3.3.

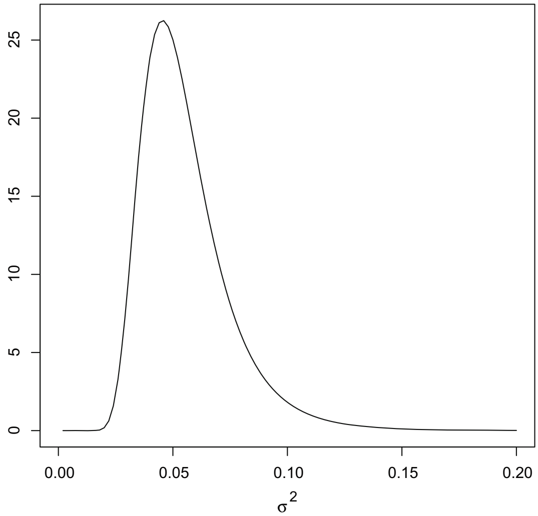

Fig. 3.3 Inverse Gamma prior distribution $\mathrm{IG}(\frac{20}{2}, \frac{1}{2})$ for σ^2 in Example 3.6

Example 3.6 (Printed Documents—Continued) Consider again Example 3.5 where magnetic flux was measured on uncontested and questioned pages. The population variance σ^2 was assumed known and equal to 0.0576. Suppose now that a new measuring device is used and that the number of previous experiments (i.e., measurements) conducted with this device is limited. A conjugate prior distribution as in (3.12) is introduced to model prior uncertainty about θ and σ^2.

The prior distribution for $\theta \mid \sigma^2$ can be centered at $\mu = 17.5$ as in Example 3.5 with $n_0 = 0.004$ reflecting a very weak prior belief with respect to the precision of the observations, $\theta \sim \text{N}(17.5, \sigma^2/0.004)$.

```
> mu=17.5
> n0=0.004
```

The prior distribution about σ^2 has been elicited above, with $k = 20$ degrees of freedom, and $S = 1$, $\sigma^2 \sim \text{IG}(\frac{20}{2}, \frac{1}{2})$, shown in Fig. 3.3.

```
> library(extraDistr)
> S=1
> k=20
> plot(function(x) dinvgamma(x,k/2,S/2),0,0.2,
+ xlab=expression(paste(sigma)^2),ylab='')
```

Note that the function `dinvgamma` is available in the package `extraDistr` (Wolodzko, 2020). Measurements are the same as in Example 3.5.

```
> x=c(16,15,15,16,15,16)
> y=c(15,16,16)
> n=length(x)
> ny=length(y)
```

Let us first consider the marginal density in the numerator of the Bayes factor in (3.13). It is a Student t distribution with $2\alpha + n = k + n = 26$ degrees of freedom, centered at $\mu_n = 15.5$ with spread parameter $s_n = 20.6724$.

```
> mun=(n*mean(x)+n0*mu)/(n+n0)
> mun

[1] 15.50133

> s2=sum((x-mean(x))^2)
> bn=S/2+(s2+n0*n*(mean(x)-mu)^2*(n0+n)^(-1))/2
> sn=ny*(n+n0)/(n+n0+ny)*(k+n)/2*bn^(-1)
> sn

[1] 20.6724
```

(continued)

Example 3.6 (continued)
The marginal density at the denominator of the Bayes factor in (3.13) is a Student t distribution with $2\alpha = k = 20$ degrees of freedom, centered at $\mu = 17.5$ with spread parameter $s_d = 0.0799$.

```
> sd=ny*n0/(n0+ny)*k/S
> sd
```

```
[1] 0.07989348
```

The density of a non-central Student t distributed random variable can be calculated using the function `dstp`, available in the package `LaplacesDemon` (Hall et al., 2020). The Bayes factor can be obtained as

```
> library(LaplacesDemon)
> BF=dstp(mean(y),mun,sn,k+n)/dstp(mean(y),mu,sd,k)
> BF
```

```
[1] 13.88188
```

The Bayes factor represents moderate support for the proposition according to which page two has been printed by the same device as the one used for printing pages one and three, compared to the proposition according to which page two has been printed by a different device.

It is worth emphasizing that the BF is highly sensitive to the choice of the prior (see Sect. 1.11). A sensitivity analysis should therefore be conducted.

3.3.3 Normal Model for Inference of Source

Consider again a case as described in Sect. 3.3.1, involving the analysis of toner on printed documents. Magnetic flux was considered as a feature of interest because it is largely influenced by the settings of the printing device. Suppose now that more than one potential source (i.e., printing device) is available for examination. The issue of interest is which of two machines has been used to print a questioned document (e.g., a contested contract). The propositions of interest can be defined as follows:

H_1 : The questioned document has been printed with machine A.
H_2 : The questioned document has been printed with machine B.

The two potential sources, i.e., machines A and B, are used to print documents under controlled conditions. The measurements made on documents printed by the two devices are denoted $\{\mathbf{x}_p\} = (\mathbf{x}_{pi},\ \ p = A, B \text{ and } i = 1, \ldots, m)$, with

$\mathbf{x}_{pi} = (x_{pi1}, \ldots, x_{pin})$ denoting the vector of n measurements for each analyzed page, $i = 1, \ldots, m$, from each printer $p = A, B$. Measurements are assumed to be normally distributed with unknown mean θ_p, $p = A, B$, and variance σ^2. The variance is assumed to be known and equal for the two devices. A conjugate normal prior distribution is taken for the unknown mean θ_p, say $\theta_p \sim \mathrm{N}(\mu_p, \tau_p^2)$, $p = A, B$.

Measurements on the questioned document are denoted by $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_q)$, with $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})$ denoting the vector of n measurements from each contested page $j = 1, \ldots, q$. For cases in which $q > 1$, it is assumed that all pages have been printed with a single device. The distribution of measurements on the questioned document is also taken to be normal. The sample mean $\bar{y} = \frac{1}{nq}\sum_{j=1}^{q}\sum_{k=1}^{n} y_{jk}$ has a normal distribution with mean θ_p and variance σ^2/nq, $(\bar{Y} \mid \theta_p, \sigma^2/nq) \sim \mathrm{N}(\theta_p, \sigma^2/nq)$.

The Bayes factor can be computed as

$$\begin{aligned} \mathrm{BF} &= \frac{\int f(\bar{y} \mid \theta_A) f(\theta_A \mid \mathbf{x}_A) d\theta_A}{\int f(\bar{y} \mid \theta_B) f(\theta_B \mid \mathbf{x}_B) d\theta_B} \\ &= \frac{f(\bar{y} \mid \mathbf{x}_A, H_1)}{f(\bar{y} \mid \mathbf{x}_B, H_2)}. \end{aligned} \tag{3.14}$$

The marginal probability density in the numerator can be obtained in closed form. It is a normal distribution with mean equal to the posterior mean $\mu_{A,x}$ and variance equal to the sum of the posterior variance $\tau_{A,x}^2$ and population variance σ_A^2/nq (where nq is the total number of observations), that is, $f(\bar{y} \mid \mathbf{x}_A, H_1) = \mathrm{N}(\mu_{A,x}, \tau_{A,x}^2 + \sigma^2/nq)$. In the same way, one can obtain the marginal probability density in the denominator, $f(\bar{y} \mid \mathbf{x}_B, H_2) = \mathrm{N}(\mu_{B,x}, \tau_{B,x}^2 + \sigma^2/nq)$. As observed in Sect. 3.3.1, the numerator and the denominator of (3.14) can be calculated as the densities of two normally distributed random variables, $\mathrm{N}(\mu_{A,x}, \tau_{A,x}^2 + \sigma^2/nq)$ and $\mathrm{N}(\mu_{B,x}, \tau_{B,x}^2 + \sigma^2/nq)$, at the sample mean $\bar{y}$ of the measurements on the questioned document.

Example 3.7 (Printed Documents) Consider a type of case and propositions as introduced above, and suppose that there is only one contested page, that is, $q = 1$. Measurements of the magnetic flux lead to the following results: $\mathbf{y} = (20, 20, 21)$ (i.e., $n = 3$ measurements are taken). Two pages are printed with each printing device. The results are as follows (Biedermann et al., 2016a):

	Printer A	Printer B
Page 1	20 20 19	21 20 21
Page 2	20 21 20	21 22 21

(continued)

Example 3.7 (continued)
The available data thus are

```
> xa=c(20,20,19,20,21,20)
> xb=c(21,20,21,21,22,21)
> y=c(20,20,21)
> n=length(y)
```

The population standard deviation σ is taken to be equal to 0.24, as in Example 3.5. We also choose the same prior distribution as used in Example 3.5 to describe uncertainty about the magnetic flux of toner printed by the two printing devices. Thus, $\mu_A = \mu_B = 17.5$ and $\tau_A^2 = \tau_B^2 = 3.92^2$.

```
> sigma2=0.24^2
> na=length(xa)
> nb=length(xb)
> mu=17.5
> tau2=3.92^2
```

The posterior distributions $f(\theta_A \mid \mathbf{x}_A)$ and $f(\theta_B \mid \mathbf{x}_B)$ can be obtained by a single application of Bayes theorem using the full set of available measurements for each printer. The posterior parameters $\mu_{A,x}$, $\mu_{B,x}$, $\tau^2_{A,x}$ and $\tau^2_{B,x}$ can be calculated using the function `post_distr`:

```
> muapost=post_distr(sigma2,na,mean(xa),mu,tau2)[1]
> tauapost=post_distr(sigma2,na,mean(xa),mu,tau2)[2]
> mubpost=post_distr(sigma2,nb,mean(xb),mu,tau2)[1]
> taubpost=post_distr(sigma2,nb,mean(xb),mu,tau2)[2]
```

The two marginal densities in the numerator and denominator of the BF in (3.14) can be calculated at the observed value $\bar{y}$. The BF can thus be computed as the ratio of two marginal densities:

```
> BF=dnorm(mean(y),muapost,sqrt(sigma2/n+tauapost))/
+ dnorm(mean(y),mubpost,sqrt(sigma2/n+taubpost))
> BF

[1] 304.7886
```

This value represents moderately strong support for the proposition according to which the questioned page been printed using device A, rather than using device B.

Consider a "$0-l_i$" loss function as in Table 1.4. The optimal decision is to accept the view according to which the questioned page was printed by the device A (as stated by proposition H_1), rather than by device B, whenever

$$\text{BF} > \frac{l_1/l_2}{\pi_1/\pi_2}.$$

If the odds are evens, and a symmetric loss function is felt to be appropriate, the Bayes decision is to accept the view according to which the questioned document has been printed with machine A (B) whenever the BF is greater (smaller) than 1.

When available information is limited, one may choose a non-informative prior distribution for (θ, σ^2) that can be specified as

$$f(\theta, \sigma^2) = \frac{1}{\sigma^2}. \tag{3.15}$$

In this case, the marginal distribution in the numerator of the BF is proportional to a Student t distribution with $n_A - 1$ degrees of freedom, centered at the sample mean $\bar{x}_A$ with spread parameter s_n equal to

$$s_n = \frac{n_A nq}{(n_A + nq)s_A^2},$$

where $s_A = \frac{1}{n_A - 1}\sum_{i=1}^{n_A}(x_A - \bar{x}_A)^2$, n_A is the total number of observations from device A, and nq is the total number of measurements from the q contested pages (i.e., n measurements for each contested page). This can be denoted as $f_1(\bar{y} \mid \bar{x}_A, s_n, n_A - 1)$.

Vice versa, the marginal distribution in the denominator of the BF is proportional to a Student t distribution with $n_B - 1$ degrees of freedom, centered at the sample mean $\bar{x}_B$ with spread parameter s_d equal to

$$s_d = \frac{n_B nq}{(n_B + nq)s_B^2},$$

where $s_B = \frac{1}{n_B - 1}\sum_{i=1}^{n_B}(x_B - \bar{x}_B)^2$ and n_B is the total number of observations from device B. This can be denoted as $f_2(\bar{y} \mid \bar{x}_B, s_d, n_B - 1)$.

The Bayes factor can then be obtained as

$$\text{BF} = \frac{f_1(\bar{y} \mid \bar{x}_A, s_n, n_A - 1)}{f_2(\bar{y} \mid \bar{x}_B, s_d, n_B - 1)}. \tag{3.16}$$

Example 3.8 (Printed Documents—Continued) In Example 3.7, a normal prior distribution has been used for (θ, σ^2). Consider now a non-informative prior distribution as in (3.15). In order to compute the Bayes factor, one must first obtain the spread parameters s_n and s_d under the competing propositions.

(continued)

Example 3.8 (continued)

```
> s2a=var(xa)
> sn=na*n/((na+n)*s2a)
> s2b=var(xb)
> sd=nb*n/((nb+n)*s2b)
```

Note that in this case the number of contested pages q is set equal to 1. The density of a non-central Student t distributed random variable can be obtained using the function `dstp` available in the package `LaplacesDemon` (Hall et al., 2020). The Bayes factor can be obtained as follows:

```
> library(LaplacesDemon)
> BF=dstp(mean(y),mean(xa),sn,na-1)/
+ dstp(mean(y),mean(xb),sd,nb-1)
> BF
```

```
[1] 2.197
```

The Bayes factor represents weak support for the proposition according to which the questioned document has been printed with machine A, rather than with machine B.

More Than Two Propositions

Consider now the case where more than two devices are available. As in Sect. 1.6, the question is how to evaluate measurements made on questioned and known items (i.e., documents), as the BF involves pairwise comparisons. A scaled version of the marginal likelihood may be reported as in (1.27).

Example 3.9 (Printed Documents, More Than Two Propositions) Recall Example 3.7, and assume that a third printer, machine C, is available for comparative examinations. The propositions of interest are therefore:

H_1 : The questioned document has been printed with machine A.
H_2 : The questioned document has been printed with machine B.
H_3 : The questioned document has been printed with machine C.

Two pages are printed with the additional printing device C. All results, including those from machines A and B, are as follows:

(continued)

Example 3.9 (continued)

	Printer A	Printer B	Printer C
Page 1	20 20 19	21 20 21	21 20 21
Page 2	20 21 20	21 22 21	20 21 20

Let the prior distribution describing uncertainty about the magnetic flux characterizing machine C be the same as introduced previously, that is $\mu_C = 17.5$ and $\tau_C^2 = 3.92^2$. First, the posterior distribution $f(\theta_C \mid \mathbf{x}_C)$ is calculated:

```
> xc=c(21,20,21,20,21,20)
> nc=length(xc)
> mucpost=post_distr(sigma2,nc,mean(xc),mu,tau2)[1]
> taucpost=post_distr(sigma2,nc,mean(xc),mu,tau2)[2]
```

Next, consider the marginal likelihoods of the sample mean that can be obtained as

```
> mla=dnorm(mean(y),muapost,sqrt(sigma2/n+tauapost))
> mlb=dnorm(mean(y),mubpost,sqrt(sigma2/n+taubpost))
> mlc=dnorm(mean(y),mucpost,sqrt(sigma2/n+taucpost))
```

The scaled version of the marginal likelihoods then is

```
> smla=mla/(mla+mlb+mlc)
> smlb=mlb/(mla+mlb+mlc)
> smlc=mlc/(mla+mlb+mlc)
> round(c(smla,smlb,smlc),5)
```

```
[1] 0.18593 0.00061 0.81346
```

Recall from Sect. 1.6 that this is equivalent to reporting the posterior probability of competing propositions with equal prior probabilities. Therefore, if $\Pr(H_1) = \Pr(H_2) = \Pr(H_3) = \frac{1}{3}$, then proposition H_3 has received the greatest evidential support.

Alternatively, the analyst may also consider the possibility of aggregating propositions H_1 and H_2 and consider:

H_1 : The questioned document has been printed with machine C.
$\bar{H}_1$: The questioned document has been printed with machine A or B.

Example 3.10 (Printed Documents, More Than Two Propositions—Continued) When considering a single proposition H_1 compared to a composite proposition $\bar{H}_1$ as defined above, the Bayes factor can be obtained as in (1.28), with $\Pr(H_1) = 1/3$ and $\Pr(\bar{H}_1) = 2/3$.

```
> p=1/3
> m1c*(1-p)/(m1a*p+m1b*p)
[1] 8.72179
```

3.3.4 Score-Based Bayes Factor

As mentioned previously in Sect. 1.5.2, it may not be possible to specify a probability model for some types of forensic evidence and data. An example was given in Sect. 3.2.3 for discrete data regarding consecutive matching striations, used to quantify the extent of agreement between marks on bullets.

Consider now a case where a saliva trace is collected at the crime scene. The salivary microbiome is analyzed as well as that of traces originating from a known source, Mr. X, with the aim of discriminating between the following competing propositions:

H_1 : The saliva trace comes from Mr. X.
H_2 : The saliva trace comes from the twin brother of Mr. X.

Note that the proposition H_2 represents an extreme case of relatedness. To investigate this type of case, consider the data collected by Scherz (2021). This longitudinal study involving 30 monozygotic twins has shown the potential of salivary microbiome profiles to discriminate between closely related individuals (Scherz et al., 2021). This may represent an alternative method when standard DNA profiling analyses yield no useful results.

In the study by Scherz (2021), four salivary samples have been collected from each participant. The first at the beginning of the study, and the others after 1, 12, and 13 months. Given the complex composition of microbiota, a distance can be calculated to compare microbiota profiles. One possibility is the Jaccard distance, obtained by dividing the number of amplicon sequence variants (AVSs) shared by the two samples by the number of distinct AVSs in the two compared samples. This measure has shown good discriminatory power. Other distances (e.g., Jensen–Shannon) can be calculated (Scherz, 2021).

The intra-individual variability was studied by comparing all four samples of each individual. The intra-pair variability was evaluated by comparing pairs of samples from related individuals (here: homozygous twins). The inter-individual variability was studied by comparing samples of unrelated individuals (Fig. 3.4).

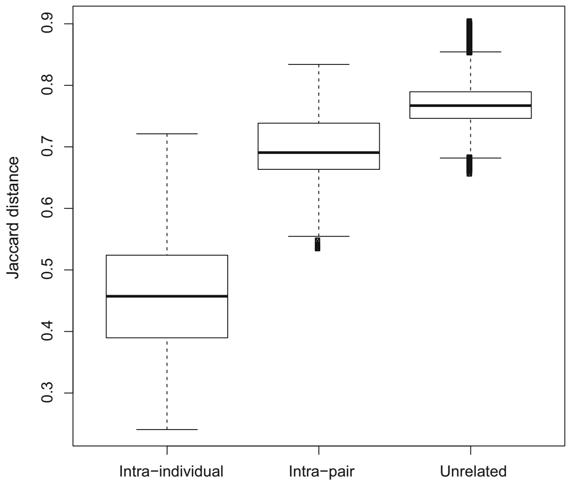

Fig. 3.4 Jaccard distances for salivary microbiota compositions of pairs of samples from individual persons (intra-individual), pairs of related persons (intra-pair), and pairs of unrelated persons (unrelated) [Source of data: (Scherz et al., 2021)]

Let $\delta(y, x)$ denote the distance between the analytical features of questioned material (i.e., a saliva trace of unknown origin) and control material (i.e., a saliva sample from Mr. X). A score-based Bayes factor (sBF) can be defined as follows:

$$\text{sBF} = \frac{g(\delta(x, y) \mid H_1)}{g(\delta(x, y) \mid H_2)}. \tag{3.17}$$

To obtain a value for this sBF, it is necessary to study the probability distribution of the calculated score under the competing propositions. However, the limited number of samples per individual, available for pairwise comparison, might make it difficult to assess the numerator, which is specific for a given person of interest. To address this problem, Davis et al. (2012) propose the use of a database of simulated samples to help with the construction of probability distributions for scores.

In the example studied here, a maximum number of 6 intra-volunteer comparisons are available for each participant. A viable alternative is to perform a so-called common-source comparison,[2] and use the limited number of items from all participants, provided that one is willing to assume a generic probability distribution for all individuals in the numerator. In the same way, a generic probability distribution is used at the denominator in all cases where a twin is assumed as the alternative source of the salivary trace (Bozza et al., 2022).

Denote by $\{Z^1_{ij},\ i = 1, \ldots, m_1,\ j = 1, \ldots, n_1\}$ the intra-individual distances and by $\{Z^2_{ij},\ i = 1, \ldots, m_2,\ j = 1, \ldots, n_2\}$ the intra-pair distances, where m_1 (m_2) are the number of distinct individuals (couples of twin brothers) and n_1 (n_2) are the number of distances calculated for each individual (couple). A normal distribution is used for both the numerator and denominator to model the *within-source* variation

[2] See Sect. 1.5.2 on the difference between specific-source and common-source propositions.

(i.e., the variation between distances characterizing materials originating from the same individual and from the same couple of twins, respectively), $Z_{ij}^p \sim \mathrm{N}(\theta_p, \sigma_p^2)$, where $p = \{1, 2\}$. Different distributions can be used to describe the between-source variation (i.e., the variation between distances characterizing materials originating from different individuals and from different couples of twins, respectively). Here, a normal distribution is retained, $\theta_p \sim \mathrm{N}(\mu_p, \tau_p^2)$. The mean vector between sources μ_p, the within-source variance σ_p^2, and the between-source variance τ_p^2 can be estimated from the background data:

$$\hat{\mu}_p = \bar{z}_p = \frac{1}{m_p n_p} \sum_{i=1}^{m_p} \sum_{j=1}^{n_p} z_{ij}^p \tag{3.18}$$

$$\hat{\sigma}_p^2 = \frac{1}{m_p(n_p - 1)} \sum_{i=1}^{m_p} \sum_{j=1}^{n_p} (z_{ij}^p - \bar{z}_i)^2 \tag{3.19}$$

$$\hat{\tau}_p^2 = \frac{1}{m_p - 1} \sum_{i=1}^{m_p} (\bar{z}_i^p - \bar{z}_p)^2 - \frac{\hat{\sigma}_p^2}{n_p}, \tag{3.20}$$

where $\bar{z}_i^p = \sum_{j=1}^{n_p} z_{ij}$.

Example 3.11 (Saliva Traces) Consider a case where a saliva trace is recovered at a crime scene and a sample is taken from a person of interest for comparative purposes. The Jaccard distance between the microbiota composition of recovered and control sample is equal to 0.51.

```
> d=0.51
```

The propositions are H_1, the compared items come from the same source, and H_2, the compared items come from different sources (twins). Suppose that the estimated means between sources in (3.18) are 0.454 and 0.769; the estimated within-source variances in (3.19) are 0.0057 and 0.00067; the estimated between-source variances in (3.20) are 0.0028 and 0.0024 (Source of data: Scherz (2021)).

```
> mu1=0.454
> mu2=0.769
> sigma1=0.0057
> sigma2=0.00067
> tau1=0.0028
> tau2=0.0024
```

The Bayes factor can then be obtained straightforwardly as in (3.17)

(continued)

Example 3.11 (continued)

```
> BF=dnorm(d,mu1,sqrt(tau1+sigma1))/
+ dnorm(d,mu2,sqrt(tau2+sigma2))
> BF
```

```
[1] 27766.33
```

The Bayes factor provides very strong support for the proposition that the saliva traces originate from the same individual rather than from two different individuals (twins).

Note that a higher value of the BF is expected whenever the alternative proposition H_2 involves unrelated individuals. The inspection of Fig. 3.4 highlights that higher distances are recorded in this type of case.

The between-source variability can also be modeled by a kernel density distribution, as presented in Bozza et al. (2022). See also Sect. 3.4.1.2, where a detailed description of the kernel density approach is given for two-level multivariate data.

3.4 Multivariate Data

Forensic scientists encounter multivariate data in contexts where the examined objects and materials can be described by several variables. Examples are glass fragments that are searched and recovered on the clothing of a person of interest and on a crime scene, or seized materials supposed to contain illicit substances. Such materials may be analyzed and compared on the basis of their chemical compounds as well as their physical characteristics. Multivariate data also arise in other forensic science disciplines, such as handwriting examination. Handwritten characters can, in fact, be described by means of several variables, such as the width, the height, the surface, the orientation of the strokes, or by Fourier descriptors (Marquis et al., 2005). In addition, an emerging topic that forensic document examiners nowadays encounter is handwriting (e.g., signatures) on digital tablets. Such electronic devices provide several static (e.g., length of a signature) and dynamic features (e.g., speed) that can be used as variables to describe signatures (Linden et al., 2018). These developments have led to substantial databases that often present a complex dependence structure, a large number of variables, and multiple sources of variation.

3.4.1 Two-Level Models

Denote by p the number of characteristics (variables) observed on items of a particular evidential type. Suppose that continuous measurements of these variables are available on a random sample of m sources with n items from each source. For handwriting evidence, a source is a single writer, with n characters from each writer and p observed characteristics that pertain to the shape of handwritten characters. For glass evidence, a source is a window, with n replicate measurements from a glass fragment originating from each window and p observed characteristics given by concentrations in elemental composition. The background data can be denoted by $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijp})$, where $i = 1, \ldots, m$ denotes the number of sources (e.g., windows), $j = 1, \ldots, n$ denotes the number of items for each source (e.g., replicate measurements from a glass fragment), and p is the number of variables.

This data structure suggests a two-level hierarchy, accounting for two sources of variation: the variation between replicate measurements within the same source (the so-called within-source variation) and the variation between sources (the so-called between-source variation).

3.4.1.1 Normal Distribution for the Between-Source Variability

In some applications, data exhibit regularity that can reasonably be described using standard probabilistic models. For example, the within-source variability and the between-source variability may be modeled by a normal distribution. A Bayesian statistical model for the evaluation of trace evidence for two-level normally distributed multivariate data was proposed by Aitken and Lucy (2004) in the context of evaluating the elemental composition of glass fragments. To illustrate this model, denote the mean vector within source i by $\boldsymbol{\theta}_i$. Denote by W the matrix of within-source variances and covariances. The distribution of Z_{ij} for the within-source variation is taken to be normal, $Z_{ij} \sim \mathrm{N}(\boldsymbol{\theta}_i, W)$. For the between-source variation, the mean vector between sources is denoted by $\boldsymbol{\mu}$, and the matrix of between-source variances and covariances by B. The distribution of the $\boldsymbol{\theta}_i$ is taken to be normal, $\boldsymbol{\theta}_i \sim \mathrm{N}(\boldsymbol{\mu}, B)$.

Measurements are available on items from an unknown source (recovered material) as well as measurements on items from a known source (control material). The examined items may or may not come from the same source. Competing propositions may be formulated as follows:

H_1 : The recovered and the control item originate from the same source.
H_2 : The recovered and the control item originate from different sources.

Denote the measurements on recovered and control items by, respectively, $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{n_y})$ and $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{n_x})$, where $\mathbf{y}_j = (y_{j1}, \ldots, y_{jp})$, $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})$, $j = 1, \ldots, n_{y(x)}$. A Bayes factor can be derived as in (1.15):

$$\text{BF} = \frac{f(\mathbf{y}, \mathbf{x} \mid H_1)}{f(\mathbf{y}, \mathbf{x} \mid H_2)}. \tag{3.21}$$

The distribution of the measurements on the recovered and control materials is taken to be normal, with vector means $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$, and covariance matrices W_y and W_x. Thus,

$$(Y \mid \boldsymbol{\theta}_y, W_y) \sim \text{N}(\boldsymbol{\theta}_y, W_y) \quad ; \quad (X \mid \boldsymbol{\theta}_x, W_x) \sim \text{N}(\boldsymbol{\theta}_x, W_x). \tag{3.22}$$

The Bayes factor is the ratio of two probability densities of the form $f(\mathbf{y}, \mathbf{x} \mid H_i) = f_i(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, W, B)$, $i = 1, 2$. The probability density in the numerator is given by

$$f_1(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, W, B) = \int_{\boldsymbol{\theta}} f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) d\boldsymbol{\theta}, \tag{3.23}$$

where

$$f(\mathbf{y} \mid \boldsymbol{\theta}, W) = |2\pi|^{-pn_y/2} |W|^{-n_y/2} \exp\left[-\frac{1}{2} \sum_{j=1}^{n_y} (\mathbf{y}_j - \boldsymbol{\theta})' W^{-1} (\mathbf{y}_j - \boldsymbol{\theta}) \right], \tag{3.24}$$

$f(\mathbf{x} \mid \boldsymbol{\theta}, W)$ has the same probabilistic structure as $f(\mathbf{y} \mid \boldsymbol{\theta}, W)$, and

$$f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) = |2\pi|^{-p/2} |B|^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' B^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) \right]. \tag{3.25}$$

In the denominator, where $\mathbf{y}$ and $\mathbf{x}$ are taken to be independent, the probability density is given by

$$\begin{aligned} f_2(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, W, B) &= f_2(\mathbf{y} \mid \boldsymbol{\theta}, W, B) \times f_2(\mathbf{x} \mid \boldsymbol{\theta}, W, B) \qquad (3.26) \\ &= \int_{\boldsymbol{\theta}} f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) d\boldsymbol{\theta} \int_{\boldsymbol{\theta}} f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) d\boldsymbol{\theta}. \end{aligned}$$

This is equivalent to the algebraic expression of the Bayes factor in (1.23). In the numerator, under proposition H_1, the source means $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$ are assumed equal, say $\boldsymbol{\theta}_y = \boldsymbol{\theta}_x = \boldsymbol{\theta}$. In the denominator, under proposition H_2, the source means $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$ are assumed to be different.

The integrals in (3.23) and (3.26) have an analytical solution. A proof is given by Aitken and Lucy (2004). The numerator can be shown to be equal to

$$\begin{aligned} f(\mathbf{y}, \mathbf{x} \mid H_1) = &| 2\pi W |^{-(n_y+n_x)/2} | 2\pi B |^{-1/2} | 2\pi \left[(n_y + n_x) W^{-1} + B^{-1} \right]^{-1} |^{\frac{1}{2}} \\ &\times \exp\left\{ -\frac{1}{2} \left[F_1 + F_2 + \text{tr}\left(S_y W^{-1} \right) + \text{tr}\left(S_x W^{-1} \right) \right] \right\}, \qquad (3.27) \end{aligned}$$

where:

$$F_1 = (\bar{\mathbf{w}} - \boldsymbol{\mu})' \left(\frac{W}{n_y+n_x} + B\right)^{-1} (\bar{\mathbf{w}} - \boldsymbol{\mu}),$$
$$F_2 = (\bar{\mathbf{y}} - \bar{\mathbf{x}})' \left(\frac{W}{n_y} + \frac{W}{n_x}\right)^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{x}}),$$
$$\bar{\mathbf{w}} = \frac{1}{n_y+n_x} \left(\sum_{j=1}^{n_y} \mathbf{y}_j + \sum_{j=1}^{n_x} \mathbf{x}_j\right), \bar{\mathbf{y}} = \frac{1}{n_y} \sum_{j=1}^{n_y} \mathbf{y}_j \text{ and } \bar{\mathbf{x}} = \frac{1}{n_x} \sum_{j=1}^{n_x} \mathbf{x}_j \ ,$$
$$S_y = \sum_{j=1}^{n_y} \left(\mathbf{y}_j - \bar{\mathbf{y}}\right) \left(\mathbf{y}_j - \bar{\mathbf{y}}\right)', S_x = \sum_{j=1}^{n_x} \left(\mathbf{x}_j - \bar{\mathbf{x}}\right) \left(\mathbf{x}_j - \bar{\mathbf{x}}\right)'.$$

Consider the first factor in the denominator, $f_2(\mathbf{y} \mid \boldsymbol{\theta}, W, B)$. It can be obtained as

$$f_2(\mathbf{y} \mid \boldsymbol{\mu}, W, B) =\mid 2\pi W \mid^{-n_y/2} \mid 2\pi B \mid^{-1/2} \mid 2\pi (n_y W^{-1} + B^{-1})^{-1} \mid^{1/2} \times \exp\left\{-\frac{1}{2}\left[(\bar{\mathbf{y}} - \boldsymbol{\mu})'(n_y^{-1} W + B)^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}) + \text{tr}\left(S_y W^{-1}\right)\right]\right\}. \quad (3.28)$$

The second factor $f_2(\mathbf{x} \mid \boldsymbol{\theta}, W, B)$ can be obtained analogously as

$$f_2(\mathbf{x} \mid \boldsymbol{\mu}, W, B) =\mid 2\pi W \mid^{-n_x/2} \mid 2\pi B \mid^{-1/2} \mid 2\pi (n_x W^{-1} + B^{-1})^{-1} \mid^{1/2} \times \exp\left\{-\frac{1}{2}\left[(\bar{\mathbf{x}} - \boldsymbol{\mu})'(n_x^{-1} W + B)^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) + \text{tr}\left(S_x W^{-1}\right)\right]\right\}. \quad (3.29)$$

The Bayes factor in (3.21) then is the ratio between (3.27) and the product between (3.28) and (3.29), respectively. After some manipulation, the BF can be obtained as the ratio between

$$\mid 2\pi \left[(n_y + n_x) W^{-1} + B^{-1}\right]^{-1} \mid^{1/2} \exp\left\{-\frac{1}{2} (F_1 + F_2)\right\} \quad (3.30)$$

and

$$\mid 2\pi B \mid^{-1/2} \mid 2\pi (n_y W^{-1} + B^{-1})^{-1} \mid^{1/2} \mid 2\pi (n_x W^{-1} + B^{-1})^{-1} \mid^{1/2} \times \exp\left\{-\frac{1}{2} (F_3 + F_4)\right\}, \quad (3.31)$$

where:

$$F_3 = (\boldsymbol{\mu} - \boldsymbol{\mu}^*)' \left\{\left(\frac{W}{n_y} + B\right)^{-1} + \left(\frac{W}{n_x} + B\right)^{-1}\right\} (\boldsymbol{\mu} - \boldsymbol{\mu}^*),$$
$$F_4 = (\bar{\mathbf{y}} - \bar{\mathbf{x}})' \left(\frac{W}{n_y} + \frac{W}{n_x} + 2B\right)^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{x}}),$$
$$\boldsymbol{\mu}^* = \left\{\left(\frac{W}{n_y} + B\right)^{-1} + \left(\frac{W}{n_x} + B\right)^{-1}\right\}^{-1} \times \left\{\left(\frac{W}{n_y} + B\right)^{-1} \bar{\mathbf{y}} + \left(\frac{W}{n_x} + B\right)^{-1} \bar{\mathbf{x}}\right\}.$$

The mean vector between sources $\boldsymbol{\mu}$, the within-source covariance matrix W, and the between-source covariance matrix B can be estimated using the available background data:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{z}} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{z}_{ij}, \tag{3.32}$$

$$\hat{W} = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{z}_{ij} - \bar{\mathbf{z}}_i)(\mathbf{z}_{ij} - \bar{\mathbf{z}}_i)', \tag{3.33}$$

$$\hat{B} = \frac{1}{m-1} \sum_{i=1}^{m} (\bar{\mathbf{z}}_i - \bar{\mathbf{z}})(\bar{\mathbf{z}}_i - \bar{\mathbf{z}})' - \frac{\hat{W}}{n}, \tag{3.34}$$

where $\bar{\mathbf{z}}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{z}_{ij}$.

Example 3.12 (Glass Evidence) Consider a case in which two glass fragments are recovered on the jacket of an individual who is suspected to be involved in a crime. Two glass fragments are collected at the crime scene for comparative purposes. The competing propositions are:

H_1 : The recovered and known glass fragments originate from the same source (broken window at the crime scene).
H_2 : The recovered and known glass fragments originate from different sources.

For each fragment, three variables are considered: the logarithmic transformation of the ratios Ca/K, Ca/Si, and Ca/Fe (Aitken and Lucy, 2004). Two replicate measurements are available for each fragment. Measurements on the two recovered fragments are

$$\mathbf{y}_1 = \begin{pmatrix} 3.77379 \\ -0.89063 \\ 2.62038 \end{pmatrix}, \ \mathbf{y}_2 = \begin{pmatrix} 3.93937 \\ -0.89343 \\ 2.63860 \end{pmatrix}.$$

Measurements on the two control fragments are

$$\mathbf{x}_1 = \begin{pmatrix} 3.84396 \\ -0.91010 \\ 2.65437 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 3.72493 \\ -0.89811 \\ 2.61933 \end{pmatrix}.$$

Consider the database named `glass-data.txt`. This database is part of the supplementary material of Aitken and Lucy (2004) and contains $n = 5$ replicate measurements of the elemental concentration of glass fragments

(continued)

Example 3.12 (continued)
from several windows ($m = 62$). The variables of interest (i.e., the logarithmic transformation of the ratios Ca/K, Ca/Si, and Ca/Fe) are displayed in columns 6, 7 and 8, while the object (window) identifier is in column 9.

```
> population=read.table("glass-data.txt", header=T)
> variables=c(6,7,8)
> grouping.item=9
```

Measurements from the recovered fragments, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$, and measurements from the control fragments, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, were selected from the available replicate measurements for the first group (window). The first two replicate measurements were selected to act as recovered data, while the last two replicate measurements were selected to act as control data

```
> item=1
> recovered=population[which(population[,grouping.
+ item]==item),][1:2,variables]
> recovered

   logCaK  logCaSi logCaFe
1 3.77379 -0.89063 2.62038
2 3.93937 -0.89343 2.63860

> control=population[which(population[,grouping.
+ item]==item),][4:5,variables]
> control

   logCaK  logCaSi logCaFe
4 3.72493 -0.89811 2.61933
5 3.66573 -0.89693 2.76393
```

Data concerning measurements from the first window were then excluded from the database

```
> pop.back <- population[-which(population[,grouping.
+ item]==item),]
```

The database named `pop.back` will serve as background data and can be used to estimate the model parameters $\boldsymbol{\mu}$, W and B as in (3.32), (3.33), and (3.34) by means of the function `two.level.mv.WB` contained in the routines file `two_level_functions.r`. This file is part of the supplementary materials available on the website of this book (on

(continued)

Example 3.12 (continued)
`http://link.springer.com/`) and can be run in the `R` console by inserting the command

```
> source('two_level_functions.r')
```

The mean vector between sources, the within-source covariance matrix, and the between-source covariance matrix can therefore be obtained as follows:

```
> WB <- two.level.mv.WB(pop.back,variables,
+ grouping.item)
> mu <- WB$all.means
> W <- WB$W
> B <- WB$B
> mu

      logCaK    logCaSi  logCaFe
[1,] 4.20495 -0.7425402 2.770238

> W

               logCaK       logCaSi       logCaFe
logCaK  1.688046e-02 2.792714e-05 2.783344e-04
logCaSi 2.792714e-05 6.545540e-05 8.362677e-06
logCaFe 2.783344e-04 8.362677e-06 1.294188e-03

> B

              logCaK       logCaSi       logCaFe
logCaK   0.71485025  0.099343866 -0.047824106
logCaSi  0.09934387  0.062724678 -0.007360187
logCaFe -0.04782411 -0.007360187  0.102438334
```

The Bayes factor can be calculated as the ratio between (3.27) and (3.28) using the function `two.level.mvn.BF` available in the routines file `two_level_functions.r`. This function is part of the supplementary materials available on the website of this book (on `http://link.springer.com/`). First, it is necessary to calculate the sample means $\bar{\mathbf{y}}$ and $\bar{\mathbf{x}}$ and to determine the sample size n_y and n_x

```
> ybar=as.vector(colMeans(recovered))
> xbar=as.vector(colMeans(control))
> ny=dim(recovered)[1]
> nx=dim(control)[1]
```

(continued)

Example 3.12 (continued)
The Bayes factor can be obtained as

```
> BF=two.level.mvn.BF(W, B, mu, xbar, ybar, nx, ny)
> BF
```

```
[1] 157.6265
```

This Bayes factor represents moderately strong support for the proposition according to which the recovered and the control fragments originate from the same source, rather than from different sources. This is expected because the compared measurements refer to the same fragment.

3.4.1.2 Non-normal Distribution for the Between-Source Variability

The two-level random effect model presented in the previous section is based on the assumption of normality of the between-source variability. However, in many practical applications, observations or measurements do not exhibit (enough) regularity for standard parametric models to be used. For example, a multivariate normal distribution for the mean vector $\boldsymbol{\theta}$ may be difficult to justify. It can be replaced by a kernel density estimate, which is sensitive to multimodality and skewness, and which may provide a better representation of the available data.

Starting from a database $\{\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ij1});\ i = 1, \ldots, m \text{ and } j = 1, \ldots, n)\}$, the estimate of the probability density distribution for the between-source variability can be obtained as follows:

$$f(\boldsymbol{\theta} \mid \bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_m, B, h) = \frac{1}{m} \sum_{i=1}^{m} K(\boldsymbol{\theta} \mid \bar{\mathbf{z}}_i, B, h), \tag{3.35}$$

where the kernel density function $K(\boldsymbol{\theta} \mid \bar{\mathbf{z}}_i, B, h)$ is taken to be a multivariate normal distribution centered at the group mean $\bar{\mathbf{z}}_i$, with covariance matrix $h^2 B$. The smoothing parameter h can be estimated as

$$\hat{h} = \left(\frac{4}{2p+1} \right)^{\frac{1}{p+4}} m^{-1/(p+4)}. \tag{3.36}$$

See also Silverman (1986) and Scott (1992).

We first write a function hopt that computes the estimate of the smoothing parameter.

```
> hopt=function(p,m){
+ h=(4/(2*p+1))^(1/(p+4))*m^(-1/(p+4))
+ return(h)}
```

Thus, if the number p of variables is set equal to 4 and the number of sources m is set equal to 30, the smoothing parameter h can be estimated as in (3.36)

```
> p=4
> m=30
> hopt(p,m)

[1] 0.5906593
```

The BF can be obtained as in (3.21), where a multivariate normal distribution is used for the control and the recovered measurements as in (3.22), and a kernel distribution for the between-source variability, as in (3.35). The numerator and the denominator of the BF, $f_1(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, W, B)$ and $f_2(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, W, B)$, can be obtained analytically (Aitken and Lucy, 2004). The BF is the ratio between

$$| B |^{1/2} mh^p | n_y W^{-1} + n_x W^{-1} + (h^2 B)^{-1} |^{-1/2} \exp\left\{-\frac{1}{2}F_2\right\} \sum_{i=1}^{m} \exp\left\{-\frac{1}{2}F_i\right\} \tag{3.37}$$

and

$$| n_y W^{-1} + (h^2 B)^{-1} |^{-1/2} \sum_{i=1}^{m} \exp\left\{-\frac{1}{2}F_{yi}\right\} \times | n_x W^{-1} + (h^2 B)^{-1} |^{-1/2} \sum_{i=1}^{m} \exp\left\{-\frac{1}{2}F_{xi}\right\}, \tag{3.38}$$

where:

$$F_i = (\mathbf{w}^* - \bar{\mathbf{z}}_i)' \left\{\left(n_y W^{-1} + n_x W^{-1}\right)^{-1} + \left(h^2 B\right)\right\}^{-1} (\mathbf{w}^* - \bar{\mathbf{z}}_i),$$

$$\mathbf{w}^* = \left(n_y W^{-1} + n_x W^{-1}\right)^{-1} \left(n_y W^{-1}\bar{\mathbf{y}} + n_x W^{-1}\bar{\mathbf{x}}\right),$$

$$F_{yi} = (\bar{\mathbf{y}} - \bar{\mathbf{z}}_i)' \left(\frac{W}{n_y} + h^2 B\right)^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{z}}_i),$$

$$F_{xi} = (\bar{\mathbf{x}} - \bar{\mathbf{z}}_i)' \left(\frac{W}{n_x} + h^2 B\right)^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{z}}_i).$$

Example 3.13 (Glass Evidence—Continued) Consider the case examined in Example 3.12, and suppose a kernel distribution is used to model the between-source variability (Aitken and Lucy, 2004). Start from the same database, `glass-data.txt`, covering n replicate measurements of p variables for each of $m = 62$ different sources. The smoothing parameter can be estimated using the function `hopt`, for $p = 3$.

```
> p=3
> m=62
> h=hopt(p,m)
> h
```

```
[1] 0.5119462
```

First, the group means $\bar{\mathbf{z}}_i$ must be obtained. They are an output of the function `two.level.mv.WB`, previously used to estimate the model parameters.

```
> group.means=WB$group.means
```

Here we show only the first six rows of the $(m \times p)$ matrix, where each row represents the means of the measurements $\bar{\mathbf{z}}_i = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_{ij}$.

```
> head(group.means)
```

```
    logCaK   logCaSi  logCaFe
2 4.895500 -0.346682 2.445828
3 2.581000 -0.890684 2.922228
4 4.092612 -0.801742 2.761072
5 4.290912 -0.267606 2.665930
6 4.594812 -0.405718 2.674566
7 2.543280 -0.893428 2.898054
```

The Bayes factor can then be calculated as the ratio between (3.37) and (3.38) using the function `two.level.mvk.BF` contained in the routines file `two_level_functions`. This function is part of the supplementary materials available on the website of this book (on `http://link.springer.com/`).

```
> source('two_level_functions.r')
> BF=two.level.mvk.BF(xbar,ybar,nx,ny,W,B, group.
  means, h)
> BF
```

```
[1] 151.6001
```

The Bayes factor represents moderately strong support for the proposition according to which the recovered and the control fragments originate from the same source, rather than from different sources.

A detailed comparison and discussion of the performance of these two multivariate random effect models can be found in Aitken and Lucy (2004). An alternative approach to the kernel density estimation is presented by Franco-Pedroso et al. (2016), modeling the between-source distribution by means of a Gaussian mixture model.

Note that a third level of variability could be considered. In fact, one may wish to model separately the variability between replicate measurements from a given item originating from a given source (e.g., replicate measurements from a glass fragment originating from a given window) and the variability between different items originating from a given source (e.g., different glass fragments originating from the same window). This aspect will be tackled in Sect. 3.4.4 where *three-level* models will be introduced.

3.4.1.3 Non-constant Within-Source Variability

The two-level random effect models presented in Sects. 3.4.1.1 and 3.4.1.2 are characterized by the assumption of a constant within-source variability. In other words, it was assumed that every single source has the same intra-variability. While for some type of trace evidence this assumption is acceptable (e.g., for measurements of the elemental composition of glass fragments), a constant within-source variation may be more difficult to justify in other forensic domains. Consider, for example, the case of handwriting on questioned documents where it is largely recognized that intra-variability may vary between writers (Marquis et al., 2006).

Suppose that a handwritten document of unknown source is available for comparative examinations. Handwritten items from a person who is suspected to be the writer are collected and analyzed. Multiple characters are analyzed on the questioned document and on the known writings of the person of interest. The following propositions are defined:

H_1: The person of interest wrote the questioned document.
H_2: An unknown person wrote the questioned document.

The distribution of the vector of means within group (source) θ_i is treated as explained in Sect. 3.4.1.1, i.e., $(\boldsymbol{\theta}_i \mid \boldsymbol{\mu}, B) \sim \mathrm{N}(\boldsymbol{\mu}, B)$. An inverse Wishart distribution is chosen to model the uncertainty about the within-group covariance matrix,

$$(W_i \mid \Omega, \nu) \sim W^{-1}(\Omega, \nu), \tag{3.39}$$

where Ω is the scale matrix and ν are the degrees of freedom (Bozza et al., 2008). The scale matrix Ω is elicited in a way such that the prior mean of W_i is taken to be equal to the within-group covariance matrix estimated from the available background data as in (3.33), while $\boldsymbol{\mu}$ is estimated as in (3.32) and the between-group covariance matrix is estimated as

$$\hat{B} = \frac{1}{m-1}\sum_{i=1}^{m} n(\bar{\mathbf{z}}_i - \bar{\mathbf{z}})(\bar{\mathbf{z}}_i - \bar{\mathbf{z}})'.$$

A two-level multivariate random effect model with an inverse Wishart distribution, modeling the uncertainty about the within-source covariance matrix, has also been proposed by Ommen et al. (2017).

First, consider the numerator of the Bayes factor in (3.21). If proposition H_1 holds, then $\boldsymbol{\theta}_y = \boldsymbol{\theta}_x = \boldsymbol{\theta}$ and $W_y = W_x = W$, and the marginal likelihood is as follows:

$$\begin{aligned} f(\mathbf{y}, \mathbf{x} \mid H_1) &= f_1(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, B, \Omega, \nu) \\ &= \int f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) f(W \mid \Omega, \nu) d(\boldsymbol{\theta}, W), \end{aligned} \tag{3.40}$$

where $f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B)$ is as in (3.25), and

$$f(W \mid \Omega, \nu) = \frac{c \mid \Omega \mid^{\nu-p-1} /2}{\mid W \mid^{\nu/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(W^{-1}\Omega)\right\},$$

where c is the normalizing constant (e.g., Press, 2005).

If proposition H_2 holds, then $\boldsymbol{\theta}_y \neq \boldsymbol{\theta}_x$ and $W_y \neq W_x$, and the marginal likelihood takes the following form:

$$\begin{aligned} f(\mathbf{y}, \mathbf{x} \mid H_2) &= f_2(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}, B, \Omega, \nu) \\ &= \int f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta}, W \mid \boldsymbol{\mu}, B, \Omega, \nu) d(\boldsymbol{\theta}, W) \\ &\quad \times \int f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta}, W \mid \boldsymbol{\mu}, B, \Omega, \nu) d(\boldsymbol{\theta}, W). \end{aligned} \tag{3.41}$$

The Bayes factor is the ratio between the marginal likelihoods in (3.40) and (3.41). However, these distributions are not available in closed form as the integrals do not have an analytical solution. Several approaches are available to deal with this problem. Chib (1995) estimates the marginal likelihood $f(\mathbf{y}, \mathbf{x} \mid H_i)$ by a direct application of Bayes theorem, since the marginal likelihood can be seen as the normalizing constant of the posterior density $f(\boldsymbol{\theta}, W \mid \mathbf{y}, \mathbf{x}, H_i)$. The marginal likelihood can therefore be obtained as

$$f(\mathbf{y}, \mathbf{x} \mid H_i) = \frac{f(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta}, W \mid H_i)}{f(\boldsymbol{\theta}, W \mid \mathbf{y}, \mathbf{x}, H_i)}. \tag{3.42}$$

While the likelihood function $f(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}, W)$ and the prior density $f(\boldsymbol{\theta}, W \mid H_i)$ can be easily evaluated at any parameter point $(\boldsymbol{\theta}^*, W^*)$, this is not the case for the posterior density $f(\boldsymbol{\theta}, W \mid \mathbf{y}, \mathbf{x}, H_i)$, which is not known in closed form. A Gibbs sampling algorithm (Sect. 1.8) can be applied to the set of the complete conditional densities $f(\boldsymbol{\theta} \mid W, \mathbf{y}, \mathbf{x}, H_i)$ and $f(W \mid \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}, H_i)$, and the posterior density $f(\boldsymbol{\theta}, W \mid \mathbf{y}, \mathbf{x}, H_i)$ can be approximated from the output of the Gibbs sampling algorithm as $\hat{f}(\boldsymbol{\theta}, W \mid \mathbf{y}, \mathbf{x}, H_i)$ (Chib, 1995; Bozza et al., 2008; Aitken et al., 2021).

The marginal likelihood in (3.42) can be estimated at a given parameter point $(\boldsymbol{\theta}^*, W^*)$ as

$$\hat{f}(\mathbf{y}, \mathbf{x} \mid H_i) = \frac{f(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\theta}^*, W^*) f(\boldsymbol{\theta}^*, W^* \mid H_i)}{f(\boldsymbol{\theta}^*, W^* \mid \mathbf{y}, \mathbf{x}, H_i)}.$$

The Bayes factor is then calculated as

$$\mathrm{BF} = \frac{\hat{f}(\mathbf{y}, \mathbf{x} \mid H_1)}{\hat{f}(\mathbf{y}, \mathbf{x} \mid H_2)}. \tag{3.43}$$

As mentioned in Sect. 1.8, many other approaches are available, and their efficiency should be studied and compared.

Example 3.14 (Handwriting Evidence) Consider a hypothetical case involving a handwritten document. Handwritten items from a person of interest are available for comparative examinations. The propositions of interest are therefore:

H_1 : The person of interest wrote the questioned document.
H_2 : An unknown person wrote the questioned document.

Suppose that $n_1 = 8$ characters of type `a` are collected from the questioned document and that $n_2 = 8$ characters of the same type are extracted from a document originating from the person of interest, taken for comparative purposes. The contour shape of loops of handwritten characters can be described using a methodology based on Fourier analysis (Marquis et al., 2005, 2006). In brief, the contour shape of each handwritten character loop can be described by means of a set of variables representing the surface and a set of harmonics. Each harmonic corresponds to a specific contribution to the shape and is defined by an amplitude and a phase, the Fourier descriptors.

Consider the database named `handwriting.txt` available on the book's website. It contains data on $p = 9$ variables (i.e., the surface, the amplitude and the phase of the first four harmonics), measured on several characters of type `a` collected from $m = 20$ writers. The variables of interest are displayed in columns 2 to 10. Column 1 contains the item (writer) identifier

(continued)

Example 3.14 (continued)

```
> population=read.table('handwriting.txt',
+ header=TRUE)
> names(population)=c('writer','A0','A1','B1','A2',
+'B2','A3','B3','A4','B4')
> variables=2:10
> grouping.item=1
```

In the current example, measurements **y** on the questioned document and measurements **x** on the control document were randomly selected from the available measurements on characters collected from a given writer (i.e., writer no. 1). Starting from a total number of, say, n available characters, $2 \times n_1$ characters have been selected: the first n_1 characters serve as recovered data, while the remaining serve as control data

```
> item=1
> base=population[which(population[,grouping.item]
+ ==item),]
> nr=dim(base)[1]
> n1=8
> recovered=as.matrix(base[1:n1,variables])
> control=as.matrix(base[(n1+1):(2*n1),variables])
```

Data concerning measurements from the selected writer were then excluded from the database

```
> pop.back=population[-which(population[,grouping.
+ item]==item),]
```

The database `pop.back` will serve as background data and can be used to estimate the model parameters as in (Bozza et al., 2008) using the function `two.level.mv.WB` available in the file `two_level_functions.r`.

```
> source('two_level_functions.r')
> WB = two.level.mv.WB(pop.back,variables,
+ grouping.item,nc=TRUE)
> mu = t(WB$all.means)
> W = WB$W
> B = WB$B
```

The number of degrees of freedom ν of the inverse Wishart distribution is chosen so as to reduce the variability of this distribution, centered at the within-source covariance matrix estimated as in (3.33).

```
> p=9
> nu=40
> Omega=W*(nu-2*p-2)
```

(continued)

Example 3.14 (continued)
The Gibbs sampling algorithm is run over 10000 iterations with a burn-in of 1000.

```
> n.iter=10000
> burn.in=1000
```

The Bayes factor in (3.43) can then be calculated using the function `two.level.mvniw.BF` that is part of the supplementary materials. Note also that this routine requires other routines that are available in the packages `MCMCpack` (Martin et al., 2021) and `mvtnorm` (Genz et al., 2020).

```
> BF=two.level.mvniw.BF(recovered,control,Omega,B,mu,
+ nu,p, n.iter,burn.in)
> BF
```

```
[1] 5543330
```

The Bayes factor represents extremely strong support for the proposition according to which the questioned and the recovered handwritten materials originate from the same source, rather than from different sources. A fully documented open-source package (Gaborini, 2019) has been developed by Gaborini (2021).

Note that it is important to critically examine large BF values, such as the one obtained above. For a discussion about extreme values, see Aitken et al. (2021), Hopwood et al. (2012), and Kaye (2009). Moreover, as underlined in Sect. 1.11, the marginal likelihood is highly sensitive to the prior assessments and so is the BF. In particular, while the overall mean vector, the within- and the between-source covariance matrices are estimated from the available background data, the number of degrees of freedom of the inverse Wishart distribution are chosen so as to reduce the dispersion of the prior. A sensitivity analysis may be performed to assess the sensitivity of the BF to different choices of the degrees of freedom ν in (3.39).

The BF may also be sensitive to the MCMC approximation. Figure 3.5 provides an illustration of BF variability. Results are based on 50 realizations of the BF approximation in (3.43).

```
> ns=50
> BFs=matrix(0,nrow=ns,ncol=1)
> for(i in 1:ns){
+ BFs[i]=two.level.mvniw.BF(recovered,control,Omega,B,
+ mu,nu,p,n.iter,burn.in)}
> hist(log(BF),freq=F,main='',xlab='log(BF)')
```

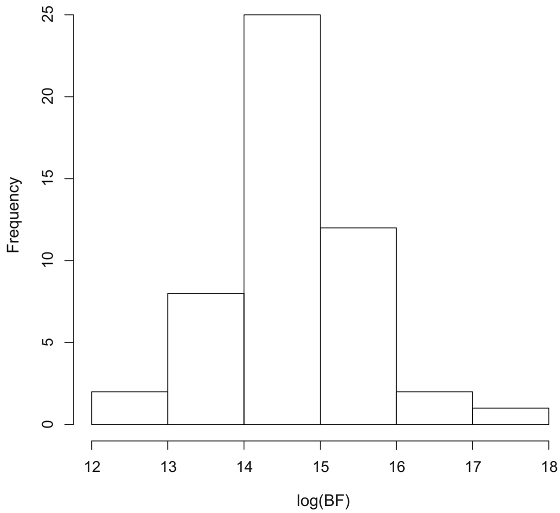

Fig. 3.5 Histogram of 50 realizations of the BF approximation in (3.43)

The models discussed here rely on the assumption of independence between sources, focusing on the inherent variability of features. In the case of questioned documents (Sect. 3.4.1.3), this amounts to assume that handwritten material has been produced without any intention of reproducing someone else's writing style. The possibility of forgery and/or disguise breaks the independence assumption made at denominator. Section 3.4.3 will address this complication.

3.4.2 Assessment of Method Performance

The results of the procedures described in the previous sections may be sensitive to changes in the features of recovered and control materials, the available background information, as well as to choices made during probabilistic modeling and prior elicitation. A sensitivity analysis may be conducted in order to gain a better understanding of the properties of the chosen method. It is fundamental to gain an understanding of how well a method performs: if the recovered and control data originate from the same source, the BF is expected to be greater than 1. Vice versa, if the compared items come from different sources, a BF smaller than 1 is expected.

Several methods exist for the assessment of the performance of the methods for evidence evaluation. Commonly encountered measures in this context are rates of false negatives (i.e., cases in which the Bayes factor is smaller than 1, supporting hypothesis H_2, when hypothesis H_1 holds) and false positives (i.e., cases in which the Bayes factor is greater than 1, supporting hypothesis H_1, when hypothesis H_2 holds). The rate of false negatives is the number of same-source comparisons with a Bayes factor smaller than 1 divided by the total number of same-source comparisons. The false positive rate is the number of different-source comparisons with a Bayes factor greater than 1 divided by the total number of different-source comparisons. Given a database of cases (e.g., measurements on handwriting characters) for which the source is known, it is possible to study the behavior of the Bayes factor as the data pertaining to control and recovered items change.

Consider again the questioned document case discussed in Sect. 3.4.1.3. There is variability in handwriting, and the reported Bayes factor is sensitive to variability of the shape of handwritten characters. This is not surprising as no one writes the same word exactly the same way twice. Consider measurements of features of handwritten characters of a given writer taken from the available database. These measurements are organized into a $(n \times p)$ matrix, where n is the number of available handwritten characters and p represents the number of features (variables). Denote this matrix `base`. Suppose that, among the n characters, we select a certain number $2 \times n_1 < n$ of characters, forming a group. Repeating this a certain number of times leads to multiple groups. On each member (character) within a group, p variables are measured. Then we take pairs of groups (i.e., measurements on the group members), taken to represent recovered and control data. Then, the Bayes factor is calculated for each couple. Here, each couple represents a same-source comparison.

Example 3.15 (Two-Level Model for Handwriting—Assessment of Model Performance) Recall Example 3.14 where a total number of 16 characters have been randomly selected from the available characters collected from a given writer (writer no. 1), extracted from the database `handwriting.txt`. A Bayes factor equal to 5543330 was obtained. If different sets of characters are extracted, the Bayes factor will be influenced (also) by the within-writer variability.

Suppose now that, for the same writer, $ns = 50$ distinct groups of characters (each of size 16) are drawn and split into groups of size 8 to act as questioned and control data. The Bayes factor is calculated for each of the 50 groups. Clearly, since the sampled measurements originate from the same writer, we expect Bayes factors greater than 1.

```
> ns=50
> n=dim(base)[1]
> n1=8
> BFs=matrix(0,nrow=ns,ncol=1)
> for (i in 1:ns){
+         ind=sample(1:n,2*n1,replace=F)
+         recovered=as.matrix(base[ind[1:n1],
+         variables])control=as.matrix(base
+         [ind[(n1+1):length(ind)],variables])
```

(continued)

Example 3.15 (continued)

```
+          BFs[i]=two.level.mvniw.BF(recovered,
+          control,Omega,
+          B,mu,nu,p,n.iter,burn.in)
+          }
```

Figure 3.6 shows a histogram of the results for the $ns = 50$ groups of sampled characters. No false negatives have been observed. The range of the BF values obtained is given here below

```
> range(BFs)
```

```
[1] 1.709027e+02 1.438262e+29
```

There is also variability between writers, as no two writers write exactly alike. Consider now measurements of features of handwritten characters from a different writer, say writer no. 6, drawn from the same database. These measurements are stored in a matrix denoted `base2`.

```
> item2=6
> base2=population[which(population[,grouping.item]==
+ item2),]
> n2=dim(base2)[1]
```

We first estimate the population parameters from the background population where both selected writers have been eliminated.

```
> pop.back=population[-which(population[,grouping
+ .item]==item|population[,grouping.item]==item2),]
> WB = two.level.mv.WB(pop.back,variables,
+ grouping.item,nc=TRUE)
> mu = t(WB$all.means)
> W = WB$W
> B = WB$B
> Omega=W*(nu-2*p-2)
```

Next, for each of the two writers, take 50 groups of characters (from `base` and `base2`). Each group contains 8 members, on each of which p features are measured. Then, take a group from each writer and form a so-called known different-source pair, and do this multiple times. These draws are taken to represent recovered and control data. Then, the Bayes factor is calculated for each couple.

```
> ns=50
> n=dim(base)[1]
> nc=dim(base2)[1]
> n1=8
```

(continued)

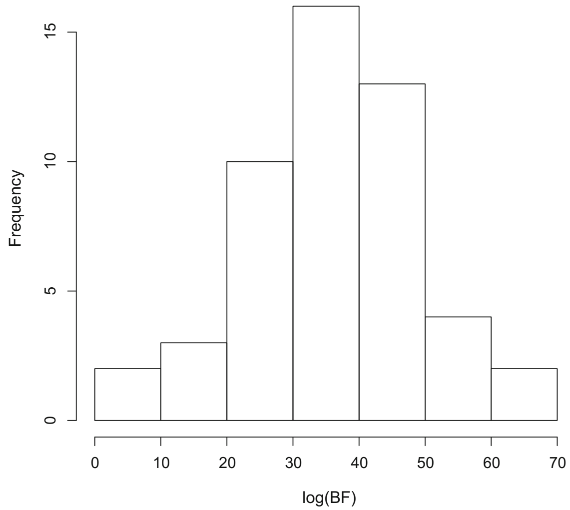

Fig. 3.6 Histogram of log(BF) values for 50 groups, each containing 8 handwritten characters, sampled from a given writer to act as questioned and control datasets

Example 3.15 (continued)

```
> BFs2=matrix(0,nrow=ns,ncol=1)
> for (i in 1:ns){
+         val.r=sample(1:n,n1)
+         recovered=as.matrix(base[val.r,variables])
+         val.c=sample(1:nc,n1)
+         control=as.matrix(base2[val.c,variables])
+         BFs[i]=two.level.mvniw.BF(recovered,
+         control,Omega,B,
+         mu,nu,p,n.iter,burn.in)
+         }
```

Figure 3.7 shows a histogram of the results. No false positives have been observed. The range of the BF values obtained is

```
> range(BFs)

[1] 2.733273e-10 7.034354e-02
```

The variability of BF values for different samples is not surprising because of handwriting variability. However, this should not be understood as there being a Bayes factor distribution. See, e.g., Morrison (2016), Ommen et al. (2016), and Taroni et al. (2016) for a discussion of issues relating to the reporting of the precision of forensic likelihood ratios.

Over the past decade, several other approaches have been proposed in forensic statistics literature for evaluating the performance of statistical procedures, based

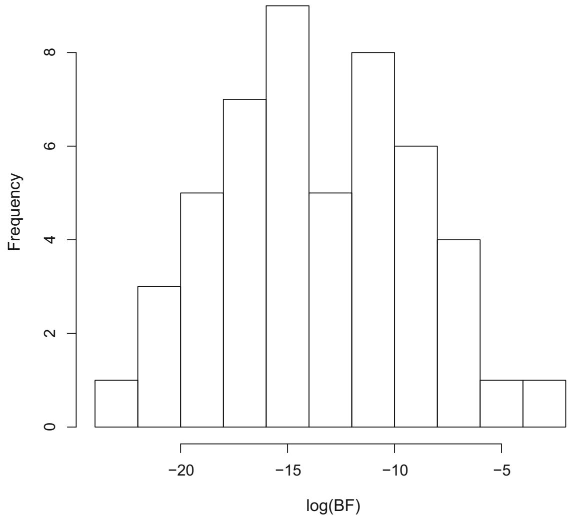

Fig. 3.7 Histogram of log(BF) values obtained for 50 groups, each containing 8 handwritten characters, sampled from the same couple of writers to act as questioned and control datasets

on a likelihood ratio or a Bayes factor. These methods provide a rigorous approach to assessing and comparing the performance of evaluative methods prior to using them in casework and forensic reporting. See, in particular, Ramos and Gonzalez-Rodriguez (2013) and Ramos et al. (2021) for a methodology to measure calibration of a set of likelihood ratio values and the concept of Empirical Cross-Entropy for representing performance, illustrated using examples from forensic speech analysis. These concepts are also discussed by Meuwly et al. (2017) who present a guideline for the validation of evaluative methods considering source level propositions. Zadora et al. (2014) present performance assessment for physicochemical data in the context of trace evidence (e.g., glass). For a recent review, see also Chapter 8 of Aitken et al. (2021).

3.4.3 On the Assumption of Independence Under H_2

The models presented in Sect. 3.4.1 are based on the assumption of independence between the questioned and known materials under hypothesis H_2. This may be reasonable for certain types of evidence and cases, but less for others. In fact, while a physical feature (e.g., the elementary composition of glass fragments) requires external constraint to be altered, a behavioral or biometric feature such as signature can be modified intentionally.

Consider handwriting as an example. When evaluating results of comparative handwriting examination, the case circumstances may be such that there is no issue of handwriting features being disguised or the result of an attempt to imitate the handwriting of another person. The approach suggested in Sect. 3.4.1.3 may thus be applicable. In turn, in case of alleged forgery of signatures, the (unknown) writer specifically intends to reproduce features of a target signature. The allegation, then, is that a signature is either simulated or disguised, rather than presenting a correspondence or similarity with a genuine signature by mere chance alone (Linden et al., 2021). In such cases, the Bayes factors previously developed in Sect. 3.4.1 cannot be used to approach the question of interest here because the assumption of independence between sources at the denominator cannot be maintained. It follows that one must compute

$$\mathrm{BF} = \frac{f(\mathbf{y} \mid \mathbf{x}, H_1)}{f(\mathbf{y} \mid \mathbf{x}, H_2)}, \tag{3.44}$$

as $f(\mathbf{y} \mid \mathbf{x}, H_2)$, following the above argument, does not simplify to $f(\mathbf{y} \mid H_2)$ (see also Sect. 1.5.1).

Consider the following competing propositions:

H_1 : The person of interest (POI) produced the questioned signature.
H_2 : An unknown person produced the questioned signature, trying to simulate the POI's signature.

If proposition H_2 is true, the forensic document examiner has to deal with a signature written by someone who has knowledge of the POI's signature.

Consider the two-level model in Sect. 3.4.1.3 where the distribution of the measurements on the recovered and control data is taken to be Normal, with vector means $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$, and covariance matrices W_y and W_x

$$(Y \mid \boldsymbol{\theta}_y, W_y) \sim \mathrm{N}(\boldsymbol{\theta}_y, W_y) \qquad ; \qquad (X \mid \boldsymbol{\theta}_x, W_x) \sim \mathrm{N}(\boldsymbol{\theta}_x, W_x). \tag{3.45}$$

The probability densities at the numerator and denominator of the BF in (3.44) can be obtained as

$$\begin{aligned} f(\mathbf{y}, \mathbf{x} \mid H_i) &= f_i(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}_i, B_i, \Omega_i, \nu_i) \\ &= \int f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta}, W \mid \mathbf{x}, \boldsymbol{\mu}_i, B_i, \Omega_i, \nu_i), \end{aligned} \tag{3.46}$$

where $(\boldsymbol{\mu}_i, B_i)$ and (Ω_i, ν_i) are the hyperparameters of the prior distributions under the competing propositions (i.e., a normal prior and an inverse Wishart prior distribution). The Bayes factor can thus be calculated as

$$\mathrm{BF} = \frac{f_1(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}_1, B_1, \Omega_1, \nu_1)}{f_2(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\mu}_2, B_2, \Omega_2, \nu_2)}. \tag{3.47}$$

Two different background databases are needed to inform model parameters under the competing propositions: a database of genuine signatures ($\mathbf{z}_{ij}$) and a database of imitated signatures ($\mathbf{s}_{ij}$). Someone who imitates a signature needs to work outside their writing habits and movement patterns. Thus, simulated signatures do not reflect the same movements and writing features as genuine signatures. Model parameter $\boldsymbol{\mu}_i$ can be estimated as in (3.32), and B_i as explained in Sect. 3.4.1.3. The scale matrix Ω_i can be chosen so as to center the prior distribution at the within-group covariance matrix W_i that can be estimated as in (3.33).

The probability densities in (3.46) are not available in closed form but can be estimated from the output of a MCMC algorithm following, for example, the ideas described in Sect. 3.4.1.3. A Gibbs sampling algorithm is implemented here. The routine is different from that developed in Sect. 3.4.1.3 because it calculates the BF in (3.47). In this formula, no assumption of independence is made at the denominator, and two different databases are used.

Example 3.16 (Digitally Captured Signatures) Consider a case involving a questioned signature on a contract signed on a digital tablet. The person of interest denies having signed the contract. Among the multiple features that are captured by the digital tablet, the average speed and writing time are considered here. See Linden et al. (2021) for a detailed description of the experimental conditions. Measurements on the questioned signature are $\mathbf{y} = (4639, 380.42)$, while measurements on the control signature are $\mathbf{x} = (4460, 323.4787)$. Note that the first value is the average speed and the second is the writing time.

```
> quest=c(4639,380.42)
> ref=c(4460,323.4787)
```

Model parameters under hypothesis H_1 (i.e., the mean vector $\boldsymbol{\mu}_1$, the within-group covariance matrix W_1, and the between-group covariance matrix B_1) are estimated from an available database of genuine signatures ($\mathbf{z}_{ij}$) and are given here below.

```
> mug=matrix(c(2754.767,511.284),ncol=1)
> Wg=matrix(c(95755.861,-4214.939,-4214.939,
+ 2857.975),byrow=T,nrow=2)
> Bg=matrix(c(3377136,30548.24,30548.24,20335.10),
+ byrow=T,nrow=2)
```

The trace matrix of the inverse Wishart distribution is then obtained as

```
> p=2
> nu=10
> Omegag=Wg*(nu-2*p-2)
```

In the same way, model parameters under hypothesis H_2 are estimated from an available database of simulated signatures ($\mathbf{s}_{ij}$) and are given here below.

(continued)

Example 3.16 (continued)

```
> mus=matrix(c(14824.3,145.0719),ncol=1)
> Ws=matrix(c(14798844,-42412.0995,-42412.0995,
+ 940.0561), byrow=T,nrow=2)
> Bs=matrix(c(37657528.8,-157142.437,-157142.437,
+ 3691.482), byrow=T,nrow=2)
> Omegas=Ws*(nu-2*p-2)
```

A Gibbs sampling algorithm is run over 10000 iterations, with a burn-in of 1000.

```
> n.iter=10000
> burn.in=1000
```

The Bayes factor in (3.44) can then be calculated using the function `two.level.mvniw2.BF` (see supplementary materials).

```
> source('two_level_functions.r')
> BF=two.level.mvniw2.BF(quest,ref,Wg,Bg,mug,Ws,Bs,
+ mus,nu,p,n.iter,burn.in)
> BF
```

```
[1] 40846.87
```

The BF represents very strong support for the proposition according to which the questioned signature originates from the person of interest rather than from an unknown person who attempted to imitate the target signature.

3.4.4 Three-Level Models

So far, two-level models have been considered, taking into account the within-source and the between-source variability. However, it is not uncommon to encounter situations in which the hierarchical ordering shows an additional level of variability, e.g., in relation to measurement error.

Denote again by p the number of variables observed on items of a given evidential type. Suppose that continuous measurements of these variables are available on a random sample from m sources with s items for each source and n replicate measurements on each of the $N = ms$ items. The background data can be denoted by $\mathbf{z}_{ikj} = (z_{ikj1}, \ldots, z_{ikjp})'$, where $i = 1, \ldots, m$ denotes the number of sources (e.g., windows, writers), $k = 1, \ldots, s$ denotes the number of items for each source (e.g., glass fragments, handwritten characters), and $j = 1, \ldots, n$ denotes the number of replicate measurements for each item.

A Bayesian statistical model for the evaluation of evidence for three-level normally distributed multivariate data was proposed by Aitken et al. (2006), focusing on the elemental composition of glass fragments. Denote the mean vector within item k in group i as $\boldsymbol{\theta}_{ik}$ and the covariance matrix of replicate measurements as W. For the variability of replicate measurements, the distribution of $\mathbf{Z}_{ikj}$ is taken to be normal, $\mathbf{Z}_{ikj} \sim \mathrm{N}(\boldsymbol{\theta}_{ik}, W)$.

Denote by $\boldsymbol{\mu}_i$ the mean vector within group i and by V the within-group covariance matrix. The distribution of $\boldsymbol{\theta}_{ik}$ for the within-group variability is taken to be normal, $\boldsymbol{\theta}_{ik} \sim \mathrm{N}(\boldsymbol{\mu}_i, B)$.

Denote by $\boldsymbol{\phi}$ the mean vector between groups. Let U denote the between-group covariance matrix. For the between-group variability, the distribution of the $\boldsymbol{\mu}_i$ is taken to be normal, $\boldsymbol{\mu}_i \sim \mathrm{N}(\boldsymbol{\phi}, V)$.

Consider the case described in Sect. 3.4.1, where measurements are available on n_y items from an unknown origin as well as measurements on n_x items from a known origin. These two groups of items may or may not come from the same source. Competing propositions may be formulated as follows:

H_1 : The recovered and the control items originate from the same source.
H_2 : The recovered and the control items originate from different sources.

There are n_1 replicate measurements available on each of the recovered n_y items. Denote the measurement vector by $\mathbf{y}$, where the vector components are denoted by $\mathbf{y}_{kj}$ (for $k = 1, \ldots, n_y$ and $j = 1, \ldots, n_1$) and $\mathbf{y}_{kj} = (y_{kj1}, \ldots, y_{kjp})'$. For each of the n_x control items, n_2 replicate measurements are available. Denote the measurement vector by $\mathbf{x}$, where the vector components are denoted ($\mathbf{x}_{kj}$, $k = 1, \ldots, n_x$ and $j = 1, \ldots, n_2$) and $\mathbf{x}_{kj} = (x_{kj1}, \ldots, x_{kjp})'$.

The Bayes factor is the ratio of two probability densities of the form $f(\mathbf{y}, \mathbf{x} \mid H_i) = f_i(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\phi}, W, B, V)$, $i = 1, 2$. The probability density in the numerator is given by

$$f_1(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\phi}, W, B, V) = \int\int f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, V) d\boldsymbol{\mu} d\boldsymbol{\theta}, \tag{3.48}$$

where all probability densities are multivariate normal.

In the denominator, the probability density is given by

$$f_2(\mathbf{y}, \mathbf{x} \mid \boldsymbol{\phi}, W, B.V) = \int\int f(\mathbf{y} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, V) d\boldsymbol{\mu} d\boldsymbol{\theta} \times \int\int f(\mathbf{x} \mid \boldsymbol{\theta}, W) f(\boldsymbol{\theta} \mid \boldsymbol{\mu}, B) f(\boldsymbol{\mu} \mid \boldsymbol{\phi}, V) d\boldsymbol{\mu} d\boldsymbol{\theta}, \tag{3.49}$$

where all probability densities are multivariate normal.

As shown by Aitken et al. (2006), the value of the evidence is the ratio of

$$\begin{aligned} &\mid B+V\mid^{1/2}\mid [(n_y n_1 + n_x n_2)W^{-1} \\ &\quad +(B+V)^{-1}]\mid^{-1/2} \exp\left\{-\frac{1}{2}(F_1+F_2)\right\} \end{aligned} \tag{3.50}$$

to

$$\begin{aligned} &\mid (n_y n_1 W^{-1} + (B+V)^{-1})\mid^{-1/2}\mid n_x n_2 W^{-1} + (B+V)^{-1}\mid^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}(F_3+F_4)\right\}, \end{aligned} \tag{3.51}$$

where:

$F_1 = (\bar{\mathbf{y}} - \bar{\mathbf{x}})'\left(\frac{n_y n_1 n_x n_2 W^{-1}}{n_y n_1 + n_x n_2}\right)(\bar{\mathbf{y}} - \bar{\mathbf{x}})$,
$F_2 = (\bar{\mathbf{w}} - \boldsymbol{\phi})'\left((n_y n_1 + n_x n_2)^{-1} W + B + V\right)^{-1}(\bar{\mathbf{w}} - \boldsymbol{\phi})$,
$F_3 = (\bar{\mathbf{y}} - \boldsymbol{\phi})'\left[(n_y n_1)^{-1} W + B + V\right]^{-1}(\bar{\mathbf{y}} - \boldsymbol{\phi})$,
$F_4 = (\bar{\mathbf{x}} - \boldsymbol{\phi})'\left[(n_x n_2)^{-1} W + B + V\right]^{-1}(\bar{\mathbf{x}} - \boldsymbol{\phi})$,

and $\bar{\mathbf{w}} = \frac{n_y n_1 \bar{\mathbf{y}} + n_x n_2 \bar{\mathbf{x}}}{n_y n_1 + n_x n_2}$.

The overall mean $\boldsymbol{\phi}$, the measurement error covariance matrix W, the within-group covariance matrix B, and the between-group covariance matrix V can be estimated using the available background data:

$$\hat{\boldsymbol{\phi}} = \frac{1}{m}\frac{1}{s}\frac{1}{n}\sum_{i=1}^{m}\sum_{k=1}^{s}\sum_{j=1}^{n}\mathbf{z}_{ikj}, \tag{3.52}$$

$$\hat{W} = \frac{1}{ms(n-1)}\sum_{i=1}^{m}\sum_{k=1}^{s}\sum_{j=1}^{n}(\mathbf{z}_{ikj} - \bar{\mathbf{z}}_{ik.})(\mathbf{z}_{ikj} - \bar{\mathbf{z}}_{ik.})', \tag{3.53}$$

$$\hat{B} = \frac{1}{m(s-1)}\sum_{i=1}^{m}\sum_{k=1}^{s}(\bar{\mathbf{z}}_{ik.} - \bar{\mathbf{z}}_{i..})(\bar{\mathbf{z}}_{ik.} - \bar{\mathbf{z}}_{i..})' - \frac{\hat{W}}{n}, \tag{3.54}$$

$$\hat{V} = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{\mathbf{z}}_{i..} - \bar{\mathbf{z}}_{...})(\bar{\mathbf{z}}_{i..} - \bar{\mathbf{z}}_{...})' - \frac{\hat{B}}{s} - \frac{\hat{W}}{sn}, \tag{3.55}$$

where $\bar{\mathbf{z}}_{ik.} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{z}_{ikj}$, $\bar{\mathbf{z}}_{i..} = \frac{1}{s}\sum_{k=1}^{s}\mathbf{z}_{ik.}$ and $\bar{\mathbf{z}}_{i...} = \frac{1}{m}\sum_{i=1}^{m}\bar{\mathbf{z}}_{i..}$.

Example 3.17 (Glass Evidence—Continued) Consider again the case described in Example 3.12 where two glass fragments are recovered on the jacket of an individual who is suspected to be involved in a crime. Two glass fragments are collected at the crime scene for comparative purposes. The competing propositions are:

(continued)

Example 3.17 (continued)

H_1 : The recovered and known glass fragments originate from the same source (e.g., a broken window).

H_2 : The recovered and known glass fragments originate from different sources.

A database named `glass-database.txt` is available as part of the supplementary material of Zadora et al. (2014). It contains measurements of the elemental concentration of glass fragments from several windows ($m = 200$). For each source, there are $s = 12$ fragments with $n = 3$ replicate measurements. For each fragment, five variables are considered: the logarithmic transformation of the ratios Na/O, Mg/O, Al/O, Si/O, Ca/O. The variables of interest are displayed in columns 3, 4, 5, 6, and 8, while the object (window) identifier is in column 1. The fragment identifier is in column 2.

```
> population=read.table('glass-database.txt',
+ header=T)
> variables=c(3,4,5,6,8)
> grouping.item=1
> grouping.fragment=2
```

Three replicate measurements are available for each fragment. Using the notation introduced above

```
> ny=2
> nx=2
> n1=3
> n2=3
```

Measurements for the recovered fragments, **y**, and measurements for the control fragments, **x**, were selected from the available data for the first and second group (window) and the first two items (fragments) from these windows. Therefore, a BF smaller than 1 is expected.

```
> recovered.item=1
> control.item=2
> base_c=population[which(population[,grouping.item]
+ ==control.item),]
> base_r=population[which(population[,grouping.item]
+ ==recovered.item),]
> recovered=base_r[which(base_r[,grouping.fragment]
+ ==1|base_r[,grouping.fragment]==2),
+ c(2,variables)]
> recovered
```

(continued)

Example 3.17 (continued)

```
  fragment  logNaO  logMgO  logAlO  logSiO
1        1 -0.6603 -1.4683 -1.4683 -0.1463
2        1 -0.6658 -1.4705 -1.4814 -0.1429
3        1 -0.6560 -1.4523 -1.4789 -0.1477
4        2 -0.6309 -1.4707 -1.5121 -0.1823
5        2 -0.6332 -1.4516 -1.4996 -0.1792
6        2 -0.6315 -1.4641 -1.4883 -0.1710
   logCaO
1 -1.1096
2 -1.1115
3 -1.1118
4 -1.1306
5 -1.1332
6 -1.1291
```

```
> control=base_c[which(base_c[,grouping.fragment]==1|
+ base_c[,grouping.fragment]==2),c(2,variables)]
> control
```

```
   fragment  logNaO  logMgO  logAlO  logSiO
13        1 -0.6231 -1.3641 -1.6540 -0.0964
14        1 -0.6122 -1.3589 -1.6622 -0.0886
15        1 -0.6108 -1.3742 -1.6935 -0.1205
16        2 -0.6135 -1.3686 -1.7202 -0.1381
17        2 -0.6205 -1.3844 -1.6831 -0.1273
18        2 -0.6204 -1.3692 -1.7269 -0.1199
    logCaO
13 -0.9993
14 -0.9836
15 -1.0524
16 -1.0830
17 -1.0721
18 -1.0392
```

Next, the means of measurements $\bar{\mathbf{y}}$, $\bar{\mathbf{x}}$, and $\bar{\mathbf{w}}$ are obtained.

```
> bary=colMeans(recovered[,-1])
> barx=colMeans(control[,-1])
> barw=colMeans(rbind(recovered,control)[,-1])
```

(continued)

Example 3.17 (continued)

Data concerning measurements from the first two windows were then excluded from the database

```
> pop.back <- population[-which(population[,
+ grouping.item]==1|population[,grouping.item]==2),]
```

The database named `pop.back` will serve as background data. It can be used to estimate the model parameters $\boldsymbol{\phi}$, W, B, and V as in (3.52), (3.53), (3.54) and (3.55) by means of the function `three.level.mv.WBV` contained in the routines file `three_level_functions.r`. This file is part of the supplementary materials available on the book's website and can be run in the `R` console with the command

```
> source('three_level_functions.r')
```

The overall mean, the measurement error covariance matrix, the within-source covariance matrix, and the between-source covariance matrix can be estimated as follows:

```
> WBV=three.level.mv.WBV(pop.back,variables,
+ grouping.item,grouping.fragment)
> psi=WBV$overall.means
> W=WBV$W
> B=WBV$B
> V=WBV$V
```

The Bayes factor can be calculated as the ratio between (3.50) and (3.51) using the function `three.level.mvn.BF` available in the routines file `three_level_functions.r`. This function is part of the supplementary materials available on the book's website.

```
> BF=three.level.mvn.BF(bary,barx,barw,ny,nx,n1,n2,
+ psi,W,B,V)
> BF
```

```
[1] 0.000083299
```

The Bayes factor represents extremely strong support for the proposition according to which the recovered and the control fragments originate from different sources, rather than from the same source.

Note that the above development does not take into account the topic of variable selection. See Aitken et al. (2006) for a proposal for dimensionality reduction based on a probabilistic structure, determined by a graphical model obtained from a scaled inverse covariance matrix.

3.5 Summary of R Functions

The R functions outlined below have been used in this chapter.

Functions Available in the Base Package

colMeans: Forms column means for numeric arrays (or data frames)

d <name of distribution>, p <name of distribution> (e.g., dpois, pnorm): Calculate the density and the cumulative probability for many parametric distributions.

More details can be found in the Help menu, help.start().

Functions Available in Other Packages

dinvgamma in package extraDistr: calculates the density of an inverse gamma distribution.

dstp in package LaplacesDemon: calculates the density of a non-central Student t distribution.

Functions Developed in the Chapter

hopt: Calculates the estimates $\hat{h}$ of the smoothing parameter h.
Usage: hopt(p,m).
Arguments: p, the number of variables: m, the number of sources.
Output: A scalar value.

poisg: Computes the density of a Poisson–gamma distribution Pg(α, β, 1) at x.
Usage: poisg(a,b,x).
Arguments: a, the shape parameter α; b, the rate parameter β; x, a scalar value x.
Output: A scalar value.

post_distr: Computes the posterior distribution $\mathrm{N}(\mu_x, \tau_x^2)$ of a normal mean θ, with $X \sim \mathrm{N}(\theta, \sigma^2)$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$.
Usage: post_distr(sigma,n,barx,pm,pv).
Arguments: sigma, the variance σ^2 of the observations; n, the number of observations; barx, the sample mean $\bar{x}$ of the observations; pm, the mean μ of the prior distribution $\mathrm{N}(\mu, \tau^2)$; pv, the variance τ^2 of the prior distribution $\mathrm{N}(\mu, \tau^2)$.
Output: A vector of values, the first is the posterior mean μ_x, the second is the posterior variance τ_x^2.

two.level.mv.WB: Computes the estimate of the overall mean $\boldsymbol{\mu}$, the group means $\bar{\mathbf{z}}_i$, the within-group covariance matrix W, and the between-group covariance matrix B for the two-level model in Sect. 3.4.1.
Usage: two.level.mv.WB(population, variables, grouping.variable,nc=FALSE).
Arguments: population, a data frame with N rows and k columns for measurements on m sources with n items for each source; variables, a vector con-

taining the column indices of the variables to be used; `grouping.variable`, a scalar specifying the variable that is to be used as the grouping factor. By default (`nc = FALSE`), the between-group covariance matrix is estimated as in Sect. 3.4.1.1. If `nc = TRUE`, the between-group covariance matrix is estimated as in Sect. 3.4.1.3.

Output: The group means $\bar{\mathbf{z}}_i$, the estimated overall mean $\hat{\boldsymbol{\mu}}$, the estimated within-group covariance matrix $\hat{W}$, the estimated between-group covariance matrix $\hat{B}$.

`two.level.mvn.BF`: Computes the BF for a two-level random effect model where both the within-source variability and the between-source variability are normally distributed, and the within-source covariance matrix is constant between sources.

Usage: `two.level.mvn.BF(W,B,mu,xbar,ybar,nx,ny)`.

Arguments: `W`, the within-source covariance matrix; `B`, the between-source covariance matrix; `mu`, the mean vector between sources; `xbar`, the vector of means for the control item; `ybar`, the vector of means for the recovered item; `nx`, the number of measurements for the control material; `ny`, the number of measurements for the recovered material.

Output: A scalar value.

`two.level.mvk.BF`: Computes the BF for a two-level random effect model where the within-source variability is normally distributed, the normal distribution for the between-source variability is replaced by a kernel density distribution, and the within-source covariance matrix is constant between sources.

Usage: `two.level.mvk.BF(xbar,ybar,nx,ny,W,B,group.means,h)`.

Arguments: `xbar`, the vector of means for the control item; `ybar`, the vector of means for the recovered item; `nx`, the number of measurements for the control material; `ny`, the number of measurements for the recovered material; `W`, the within-source covariance matrix; `B`, the between-source covariance matrix; `group.means`, a $(m \times p)$ matrix, where each row represents the vector of means $\bar{\mathbf{z}}_i = \frac{1}{n}\sum_{j=1}^{n} \mathbf{z}_{ij}$; `h`, the smoothing parameter.

Output: A scalar value.

`two.level.mvniw.BF`: Computes the BF for a two-level random effect model where both the within-source variability and the between-source variability are normally distributed, and the uncertainty about the within-source covariance matrix is modeled by an inverse Wishart distribution.

Usage: `two.level.mvniw.BF(quest,ref,O,B,mu,nw,p,n.iter, burn.in)`.

Arguments: `quest`, a $(n \times p)$ matrix containing measurements on the questioned material; `ref`, a $(n \times p)$ matrix containing measurements on the control material; `O`, the trace matrix of the inverse Wishart distribution; `B`, the between-source covariance matrix; `mu`, the mean vector between sources; `nw`, the number of degrees of freedom of the inverse Wishart distribution; `p`, the number of variables; `n.iter`, the number of iterations of the Gibbs sampling algorithm; `burn.in`, the number of discarded iterations.

Output: A scalar value.

`two.level.mvniw2.BF`: Computes the BF for a two-level random effect model where both the within-source variability and the between-source variability are normally distributed, the uncertainty about the within-source covariance matrix is modeled by an inverse Wishart distribution with no assumption of independence between questioned and known materials at the denominator (i.e., under H_2).

Usage: `two.level.mvniw2.BF(quest,ref,Og,Bg,mug,Os,Bs,mus, nu,p,n.iter,burn.in)`.

Arguments: `quest`, a $(n \times p)$ matrix containing measurements on the questioned material; `ref`, a $(n \times p)$ matrix containing measurements on the control material; `Og`, the trace matrix of the inverse Wishart distribution from the database of genuine (handwritten) material; `Bg`, the between-source covariance matrix from the database of genuine (handwritten) material; `mug`, the mean vector between sources from the database of genuine (handwritten) material; `Os`, the trace matrix of the inverse Wishart distribution from the database of simulated (handwritten) material; `Bs`, the between-source covariance matrix from the database of simulated (handwritten) material; `mus`, the mean vector between sources from the database of simulated (handwritten) material; `nw`, the number of degrees of freedom of the inverse Wishart distribution; `p`, the number of variables; `n.iter`, the number of iterations of the Gibbs sampling algorithm; `burn.in`, the number of discarded iterations.

Output: A scalar value.

`three.level.mv.WBV`: Computes the estimate of the overall mean $\boldsymbol{\phi}$, the measurement error covariance matrix W, the within-group covariance matrix B, and the between-group covariance matrix V for the three-level model presented in Sect. 3.4.4.

Usage: `three.level.mv.WBV(population,variables,grouping. item,grouping.fragment)`.

Arguments: `population`, a data frame with msn rows and k columns collecting measurements on m sources with s items for each source and n replicate measurements for each item; `variables`, a vector containing the column indices of the variables to be used; `grouping.item`, a scalar specifying the variable that is to be used as the grouping item; `grouping.fragment`, a scalar specifying the variable that is to be used for the grouping fragment.

Output: The estimated overall mean $\hat{\boldsymbol{\phi}}$, the estimated measurement error covariance matrix $\hat{W}$, the estimated within-group covariance matrix $\hat{B}$, the estimated between-group covariance matrix $\hat{V}$.

`three.level.mvn.BF`: Computes the BF for a three-level random effect model where the variation at all three levels is normally distributed.

Usage: `three.level.mvn.BF(bary,barx,barw,ny,nx,n1,n2,psi, W,B,V)`.

Arguments: `bary`, the mean vector of measurements on recovered items; `barx`, the mean vector of measurements on control items; `barw`, the mean vector of measurements; ny, the number of recovered items; nx, the number of control items; n1, the number of replicate measurements on each of the recovered items; n2, the number of replicate measurements on each of the control items; psi, the overall mean vector; W, the replicate measurements covariance matrix; B, the within-group covariance matrix; V, the between-source covariance matrix.

Output: A scalar value.

Published with the support of the Swiss National Science Foundation (Grant no. 10BP12_208532/1).