Reference Analysis in a Misclassification Model for the Meiosis I Nondisjunction Fraction in Trisomies

Rosangela H. Loschi and Vinicius D. Mayrink

Department of Statistics, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

Correspondence should be addressed to Rosangela H. Loschi; loschi@est.ufmg.br

Received 22 October 2012; Accepted 9 November 2012

Academic Editors: A. Esquela Kerscher, R. W. Pereira, and T. Saito

Copyright © 2013 R. H. Loschi and V. D. Mayrink. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The determination of the meiosis I nondisjunction fraction plays an important role in identifying the characteristics of affected individuals and their mothers, which can generate aneuploidies. The number of individuals with one, two, and three peaks pattern is used to obtain the information; however, the data are susceptible to misclassification. We review the misclassification model previously introduced in the literature which considers a common misclassification error. This paper aims to introduce a joint prior distribution for the meiosis I nondisjunction fraction and the misclassification error. We prove that the reference prior is a proper distribution. We analyze a Brazilian Down syndrome dataset and compare the results with those obtained through Bayes-Laplace and beta prior distributions.

1. Introduction

In humans, aneuploidies are common causes of mental retardation, pregnancy losses, and fetal death. Although the causes of aneuploidies are unknown, it is known that the risk of having children with some kind of aneuploidy, such as trisomies 21 (Down syndrome), 18 (Edward’s syndrome), or 13 (Patau syndrome), increases with the mother’s age [1].

Trisomy 21 is the most prevalent human genetic disorder and occurs in approximately 1 out of 700 births. It is the most common cause of mental retardation of genetic origin. Down syndrome affects the cognitive abilities of the child and approximately half of them can also have congenital heart defects and problems with hearing and vision, and they are prone to develop pulmonary hypertension. The causes of Down syndrome are unknown, but there is evidence that in the trisomy of chromosome 21 the rate of nondisjunction increases with the age of the mother [2]. Women aged 35 or older have significantly higher risk of having a child with Down syndrome. In addition, the increase in the rate of non-disjunction in meiosis II is higher than in meiosis I, if the mother is between 35 and 39 years old. As a result, the determination of the fraction $\phi$ of non-disjunction in chromosomal segregation, taking place in meiosis I in each chromosome, plays an important role in understanding meiosis I in each chromosome, plays an important role in understanding aneuploidies. It is useful to identify possible factors generating such abnormalities, for example: geography, nutrition, age, reproductive, practices. Prenatal diagnosis of aneuploidies is usually done by employing chromosome karyotype, fluorescent in situ hybridization (FISH), and polymerase chain reaction (PCR-) based approaches, see [1, 3–6] for further details.

Methods to estimate $\phi$ considering information from the affected children and their parents are presented in [7–15] among others. More recently, Bayesian and classical approaches to infer about $\phi$, assuming models that do not take into account the parental information, are presented in [2, 16].

The model proposed in [16] considers that, using the PCR, it is possible to type microsatellites located near the chromosomal centromere (to avoid problems due to recombination) through primers designed from the unique DNA sequence flanking the tandem repeat arrays, followed by quantitative analysis by computer-assisted laser densitometry [17]. Trisomic patients will display, in informative microsatellite loci, three fragment peaks of equal intensity,
two fragments at an average 2 : 1 dosage, or one individual fragment (see an example of classification in Figure 1). The relative proportion of the three cases depends on the type of non-disjunction, although the heterozygosity level is also important. For a three-allele pattern to emerge, it is necessary for the non-disjunction to occur in the first meiotic division, the mother to be heterozygous at the relevant locus, and the allele carried by the sperm to be different from the two maternal alleles. A two-allele pattern is observed either in the first or the second meiosis division, depending on the combination of the chromosomes transmitted by the parents. One peak pattern occurs if the parents are homozygous for the allele inherited, see details in [9]. A scheme for the non-disjunction in meiotic process can be found in [16].

The model proposed in [16] brings some novelty to the analysis of trisomies in the sense that the parental information is not taken into account in the modeling. The use of archive material is possible and important in the study of rare trisomies; however, since some other peaks can also be observed, as a consequence of residuals generated by the preparation of the genetic material (see Figure 1), misclassification can occur and thus the fraction $\phi$ may be poorly estimated. Extensions to the model in [16] were introduced in [18, 19] to accommodate misclassification; such misclassification models include the one indicated in [16] as a particular case. In both papers, the inference is developed under the Bayesian paradigm using beta prior distributions for the parameters $\phi$ and $\psi$.

The specification of priors makes it possible to incorporate scientific hypotheses into the analysis and, consequently, allows us to handle complex problems and situations in which little sample information is available. Reference (or noninformative) priors are frequently used to describe prior uncertainty about a parameter. There are several methods to construct non-informative priors; we consider the Jeffreys’ approach which is widely used in the literature. The idea behind the Jeffreys prior is to provide as little prior information as possible, relative to the sample information.

In this study, we review the misclassification model for trisomies presented in [18]. The main goals are to obtain the joint Jeffreys prior for the fraction $\phi$ of non-disjunction taking place in the meiosis I, to determine the misclassification error $\psi$, and to prove that the Jeffreys prior is proper. We implement a Metropolis-Hastings algorithm to sample from the posterior distributions. A case study is developed to analyze the sample of Brazilian individuals with Down syndrome reported in [16]. We compare the results obtained from the Jeffreys prior suggested here with those obtained in [18] via Bayes-Laplace priors and those using a similar approach under the model proposed in [16].

This paper is organized as follows. In Section 2, we briefly present the misclassification model for trisomies according to [18]. We build the joint Jeffreys prior for $\phi$ and $\psi$ and present some of its properties. In Section 3, we apply the proposed method to analyze a real dataset and compare the results with those obtained in the literature. Finally Section 4 shows the main conclusions.

![Figure 1](image_url)

**Figure 1:** Laser densitometry for a trisomic patient. One peak pattern in 1, two peaks pattern in 2 and 3, and three peaks in 4.

### 2. Model Description and Inference

Here, we present the misclassification model introduced in [18] and build the Jeffreys prior for the parameters indexing the model. We also review the model proposed in [16] which ignores the presence of misclassification.

In order to build the model, we assume that the hypothesis of Hardy-Weinberg equilibrium [20] has been verified for the population; thus, we can consider as known the relative frequency $p_i, i = 1, \ldots, m$, of the allele $i$ in a multiallelic locus of microsatellites.

Let $Y_l$ be the number of individuals with $l$ peaks pattern, $l = 1, \ldots, 3$, observed in a sample of $n$ trisomic individuals; define $Y = (Y_1, Y_2, Y_3)$. Denote by $\theta_l(\cdot)$ the probability of being $l, l = 1, \ldots, 3$, the true number of peaks in the microsatellite locus of interest. As proved in [16], $\theta_l(\cdot)$ depends on the fraction $\phi \in [0, 1]$ of non-disjunction taking place in Meiosis I as follows:

\[
\begin{align*}
\theta_1(\phi) &= a \phi + b (1 - \phi), \\
\theta_2(\phi) &= c \phi + d (1 - \phi), \\
\theta_3(\phi) &= e \phi,
\end{align*}
\]

where $a = \sum_{i=1}^{m} p_i^3$, $b = \sum_{i=1}^{m} p_i^2$, $c = 3 \sum_{i=1}^{m} \sum_{j=1}^{m} p_i^2 p_j$ for all $i \neq j$, $d = \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j$ for all $i \neq j$, and $e = \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} p_i p_j p_k$ for all $i \neq j \neq k$. Note that, under this notation, we have $a + c + e = (\sum_{i=1}^{m} p_i)^3 = 1$ and $b + d = (\sum_{i=1}^{m} p_i)^2 = 1$.

#### 2.1. The Misclassification Model

In order to define the misclassification model, we also consider the auxiliary random variables $X$ and $Z$ denoting the true (nonobserved) and the observed number of peaks in a trisomic individual, respectively. In addition, denote by $\pi_l(\cdot)$ the probability of
being \(l, l = 1, \ldots, 3\), the observed number of peaks in the microsatellite locus of interest.

Let \(\eta_{jk} = P(Z = j \mid X = k)\) for \(j, k = 1, 2, 3, j \neq k\), be the probability of misclassifying an individual. It follows from probability calculus that the vector of probabilities \(\pi(\phi, \eta) = \pi_{1}(\phi, \eta), \pi_{2}(\phi, \eta), \pi_{3}(\phi, \eta)\) is given by \(\pi(\phi, \eta) = \eta \theta(\phi)\), where \(\theta(\phi) = [\theta_{1}(\phi), \theta_{2}(\phi), \theta_{3}(\phi)]^{T}\) and \(\eta\) is the \((3 \times 3)\) matrix below
\[
\eta = \begin{bmatrix}
\eta_{11} & \eta_{12} & \eta_{13} \\
\eta_{21} & \eta_{22} & \eta_{23} \\
\eta_{31} & \eta_{32} & \eta_{33}
\end{bmatrix}
\] (2)

Since \(\sum_{j=1}^{3} \eta_{jk} = 1\), for each \(k = 1, 2, 3\), it can be proved that \(\sum_{j=1}^{3} \pi_j(\phi, \eta) = \sum_{j=1}^{3} \theta_j(\phi) = 1\). As a consequence of the previous assumptions, paper [18] establishes that \(Y \mid \phi, \eta \sim \text{multinomial}[n, \pi_{1}(\phi, \eta), \pi_{2}(\phi, \eta), \pi_{3}(\phi, \eta)]\) whose probability function is
\[
f(Y \mid \phi, \eta) = \frac{n!}{y_{1}! y_{2}! y_{3}!} \prod_{j=1}^{3} \left[\pi_j(\phi, \eta)\right]^{y_j},
\] (3)

with \(\sum_{j=1}^{3} y_j = n\). An identifiable model is obtained by assuming equal probabilities of misclassification, that is, \(\eta_{jk} = \begin{cases} 1 - 2\psi, & k = j, \\ \psi, & k \neq j. \end{cases}\) (4)

As a result, the likelihood function in (3) becomes
\[
f(Y \mid \phi, \eta) = \frac{n!}{y_{1}! y_{2}! y_{3}!} \prod_{j=1}^{3} \left[\left(1 - 2\psi\right) \theta_j(\phi) + \psi \left(1 - \theta_j(\phi)\right)\right]^{y_j},
\] (5)

where \(\phi \in [0, 1]\) and \(\psi \in [0, 1/4]\).

Results in [18] are determined under the Bayesian paradigm. Because some information about \(\phi\) are available in the literature for other populations, such pieces of information were used to build more informative beta prior distributions for \(\phi\). In the literature, it is assumed a uniform prior for the misclassification error and these errors are considered independent of \(\phi\) a priori. Although the beta family is very rich in form and can represent well many different prior opinions about \(\phi\), some research prefers to perform an objective analysis. In the following, we consider the Jeffreys approach to build non-informative prior distributions.

2.2. The Jeffreys Prior for \((\phi, \psi)\). In this section, we introduce the joint Jeffreys prior for the parameters \((\phi, \psi)\) induced by the model in (5). We consider that \(\psi\) is independent of \(\phi\) and obtain the marginal Jeffreys prior for \(\phi\) assuming \(\psi\) as fixed. The Jeffreys prior for \(\psi\) is determined using the same strategy. This approach to calculate the Jeffreys prior in a multiparameter scenario is quite common and avoids some misinterpretation issues related to the posterior results that usually occur whenever the Fisher information matrix is assumed to build the Jeffreys prior on a multi-parameter model.

Consider again the notation in (1); note that by fixing \(\phi\) and taking the derivative of (5) with respect to \(\psi\), it follows that the Fisher information is
\[
I(\psi) = n \left[ \frac{1 + 3\theta_1(\phi)^2}{\theta_1(\phi) [1 - 3\psi]} + \psi \right.
\]
\[
+ \frac{1 + 3\theta_2(\phi)^2}{\theta_2(\phi) [1 - 3\psi]} + \psi + \frac{1 + 3\theta_3(\phi)^2}{\theta_3(\phi) [1 - 3\psi]} + \psi \right].
\] (6)

Since our interest lies on constructing the prior distribution for \(\psi\), the terms \(\theta_l(\phi), l = 1, 2, 3\), are assumed as constants and they do not carry information about \(\psi\). As a result, the marginal Jeffreys prior for \(\psi\) is
\[
\pi(\psi) \propto \left[ \frac{1 + 3k_1}{k_1 [1 - 3\psi]} + \psi \right]
\]
\[
+ \left[ \frac{1 + 3k_2}{k_2 [1 - 3\psi]} + \psi \right] + \left[ \frac{1 + 3k_3}{k_3 [1 - 3\psi]} + \psi \right]^{1/2},
\] (7)

for all \(\psi \in [0, 1/4]\) and any constant values \(k_i \in [0, 1], i = 1, 2, 3\), such that \(\sum_{i=1}^{3} k_i = 1\). Note that if \(k_i \neq 0\) for all \(i\), we have
\[
\lim_{\psi \to 0} \frac{1 + 3k_1^2}{k_1 [1 - 3\psi]} = \frac{1 + 3k_1^2}{k_1} < \infty.
\] (8)

Therefore, \(\pi(\psi)\) is a proper distribution. If \(k_i = 0\) for some \(i\), the propriety of \(\pi(\psi)\) is also verified. As can be seen, if we assume \(k_3 = 0\), for instance, it follows that
\[
\lim_{\psi \to 0} \frac{1 + 3k_1^2}{k_1 [1 - 3\psi]} + \psi = \lim_{\psi \to 0} \frac{1}{\psi} = \infty.
\] (9)

Consequently, there is a constant \(J \in [0, 1/4]\) such that \(1 + 3k_1^2/\left(1 - 3\psi\right) + \psi < 1 < \psi^{-1}\) for all \(\psi \in [0, J]\) and \(i = 1, 2\). As a result, we have
\[
\int_{0}^{J} \pi(\psi) d\psi < \int_{0}^{J} \left(\frac{3}{\psi}\right)^{1/2} d\psi = 2(3J)^{1/2},
\] (10)

which guarantees the propriety of \(\pi(\psi)\).

Similarly, if we fix \(\psi\) and calculate the derivative of (5) with respect to \(\phi\), the Fisher information is given by
\[
I(\phi) = n [1 - 3\psi]^2
\]
\[
\times \left[ \frac{(a - b)^2}{\theta_1(\phi) [1 - 3\psi]} + \psi \right.
\]
\[
+ \frac{(c - d)^2}{\theta_2(\phi) [1 - 3\psi]} + \psi + \frac{e^2}{\theta_3(\phi) [1 - 3\psi]} + \psi \right].
\] (11)
Thus, since the terms in (11), depending of $\psi$, do not carry any information about $\phi$, the Jeiffreys prior for $\phi$ is

$$
\pi(\phi) \propto \frac{(a-b)^2}{\theta_1(\phi)\left[1-3k_4\right]+k_4} + \frac{(c-d)^2}{\theta_2(\phi)\left[1-3k_4\right]+k_4} + \frac{e^2}{\theta_3(\phi)\left[1-3k_4\right]+k_4}^{1/2},
$$

where $\phi \in [0,1]$ and $k_4 \in [0,1/4]$. Similar to what was observed for the prior distribution of $\psi$, it can be proved that the prior for $\phi$ is proper.

Since $\psi$ and $\phi$ are assumed to be independent, the joint Jeiffreys prior for $(\phi, \psi)$ is obtained through the multiplication of the expressions (7) and (12). Figure 2 presents the marginal Jeiffreys prior of $\phi$ for three different values of $\psi$, and the marginal Jeiffreys prior of $\psi$ for three different values of $\phi$. Figure 3 shows a three-dimensional surface plot representing the joint Jeiffreys prior for $(\phi, \psi)$. Note that the three priors are decreasing functions and they tend to infinity as the parameters tend to zero. However, since the $\pi(\psi)$ and $\pi(\phi)$ are both proper distributions, the joint prior for $(\phi, \psi)$ is also proper.

The joint posterior of $(\phi, \psi)$ is obtained via Bayes theorem assuming the likelihood function in (5) and the priors given in (7) and (12). The posterior distribution cannot be obtained analytically; we use a Metropolis-Hastings algorithm to sample from it.

2.3. Model in Franco et al. The model in [16] does not consider the misclassification errors that might occur when data are obtained. Conditional on $\phi$, paper [16] shows that the random vector $Y = (Y_1, Y_2, Y_3)$ has a Multinomial distribution with parameters $n, \theta_1(\phi), \theta_2(\phi), \theta_3(\phi)$, denoted by $Y \sim \text{multinomial}[n, \theta_1(\phi), \theta_2(\phi), \theta_3(\phi)]$, which has the following probability function

$$
f(Y | \phi, \eta) = \frac{n^3}{y_1! y_2! y_3!} \prod_{j=1}^{3} \left[\theta_j(\phi)\right]^{y_j},
$$

where $\theta_j(\phi)$, $j = 1, 2, 3$, are indicated in (1). Note that this model is a particular case of the model presented in (5) whenever $\psi$ is assumed to be zero. As observed for the misclassification model, the posterior distribution of $\phi$ cannot be analytically evaluated even when a Bayes-Laplace prior distribution is assumed to describe the uncertainty about $\phi$. The posterior can be approximated through numerical algorithms [21, 22] or Monte Carlo methods. In Section 3, we show results related to this model which were obtained in [19].

3. Case Study: Down Syndrome Data

Here, we explore the dataset reported in [16] consisting of a random sample of blood from 34 Brazilian individuals with trisomy of chromosome 21. In this data set, the observed numbers of patients with one, two, and three peaks are 6, 22, and 6, respectively. The hypothesis of the Hardy-Weinberg equilibrium is verified for the Brazilian population and six alleles are found with frequencies 0.12, 0.45, 0.09, 0.31, 0.01, and 0.02.

In terms of the Markov Chain Monte Carlo (MCMC) algorithm, we run a chain of size 10,000 and assume a burn-in period involving the first 5,000 iterations to guarantee
the convergence of the chain; as a result, the final sample from the posterior distribution has 5,000 observations. We assume independent uniform reference distributions to generate the candidates required in the sampling process.

Tables 1 and 2 present the posterior summaries for $\phi$ and $\psi$ assuming the misclassification model described in Section 2. We consider two non-informative priors for $(\phi, \psi)$: the joint Jeffreys prior given in Section 2 and the Bayes-Laplace prior assumed in [18]. We also compare these results with those obtained by fitting Franco et al.’s model assuming the Bayes-Laplace prior for $\phi$.

All posterior estimates for $\phi$ (Table 1) are smaller than those observed in the literature for groups of patients with Down syndrome (see [7, 10–15, 21]); the literature results indicate the mean 0.6803 and the standard deviation 0.0678. As can be seen, the estimates under the misclassification model are also smaller than those obtained for the model proposed in [16]; the maximum likelihood estimate (MLE) calculated in [16] is 0.6551 for the same data set. In the comparison between the misclassification and Franco et al’s models, we have found that the uncertainty about $\phi$ is higher for the model including misclassification. We also observe that the posterior estimates tend to be small and the variance tend to be large when we assume the Jeffreys prior. The posterior distributions of $\psi$ (Table 2) suggest that the misclassification error is very small. The posterior mean, median, mode, and the uncertainty about $\psi$ for the Jeffreys prior model are smaller than those estimates obtained via the Bayes-Laplace prior distribution.

Figure 4 presents two histograms representing the posterior distributions of $\phi$ and $\psi$ under the proposed model. In both graphs, the distribution has unique mode and an asymmetric shape. The posterior distribution of $\phi$ indicates higher probability mass for values around 0.5. Furthermore, with 95% of probability, the Meiosis I non-disjunction fraction $\phi$ belongs to the interval $[0.1157; 0.9993]$. The posterior distribution for $\psi$ puts most of its probability mass for values close to zero and with posterior probability of 97.5%, the misclassification error is smaller than 0.1614.

4. Final Comments

In this paper, we provide a reference analysis for the misclassification model introduced by [18] to describe the uncertainty about the Meiosis I non-disjunction fraction in patients.
with trisomy. The Jeffreys prior, assuming independence between the non-disjunction fraction $\phi$ and the misclassification error $\psi$, was obtained in our analysis; we have built a Metropolis-Hastings algorithm to sample from the posterior distributions. The real data application, illustrating the use of the methodology, involves a Brazilian Down syndrome data set.

In our analysis, the Jeffreys prior determines more uncertainty in the estimation of $\phi$ compared to the Bayes-Laplace prior, but its use leads to a more precise estimate of the misclassification error. In addition, we can see that, for the Brazilian population, the Down syndrome is often a consequence of a non-disjunction in meiosis I.

The use of the Jeffreys prior and other non-informative priors permit us to fairly compare Bayesian and Classical approaches of inference. It is still an open problem to find the maximum likelihood estimator under the misclassification model discussed in this paper and the other extensions discussed in [19].

Acknowledgments

The authors would like to express their gratitude to the editors for the invitation. We also thank Professor Sergio D. J. Pena and Flavia C. Parra (UFMG), for providing the dataset. R. H. Loschi has her research supported in part by the financial support from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) of the Ministry for Science and Technology of Brazil, Grants 306085/2009-7, 304505/2006-4, 472877/2006-2. V. D. Mayrink has received financial support from the Universidade Federal de Minas Gerais through the program Doutores Recém-Contratados of the PRPq (Pro-Reitoria de Pesquisa).

References
