## A Statistical Approach to Genetic Epidemiology Corrections

Andreas Ziegler and Inke R. König

We are extremely grateful to all our students and all our colleagues who pointed us to errors in our book, and we sincerely apologize for the errors.

Lübeck, April 15, 2008
Andreas Ziegler and Inke R. König

- Page xi, Section Acknowledgments: In line 5, "Genetic Epidemiolog" should read "Genetic Epidemiology".
- Page 8, Chapter Molecular Genetics: In step 7 of Meiosis "haploid status" must be replaced by "diploid status".
- Page 65 to 68, Chapter Data Quality: Numbering of algorithms is incorrect, there are no algorithms 4.2 or 4.4.
- Page 65 to 68, Chapter Data Quality: In algorithms 4.1, 4.3, and 4.5, step "(a) Set counter $C=0$ " needs to be moved in front of the "for" loop.
- Page 67, Chapter Data Quality: In line 10 from bottom "vary" should ready "very".
- Page 68, Chapter Data Quality: In algorithm 4.5, step 2 (d) should read "If $H o m_{O} \geq H o m_{P}$, add 1 to $C^{\prime \prime}$.
- Page 71, Chapter Data Quality: In example 4.6, equation after 2nd paragraph, in the middle part, it should be $0.3826^{2}$ instead of 0.3826 . In the next line, $v_{n}^{2}$ should be $\mathbb{V a r}(d)$.
- Page 72, Chapter Data Quality: In Problem 4.2, the sentence "If so, which error is most likely?" should be replaced with: "If not, which error is most likely?".
- Page 82, Chapter Genetic Map Distances: In line 17, the sentence "LD maps are based on recombination events exactly as genetic map functions." should read "LD maps are based on recombination events exactly as genetic map functions are."
- Page 94, Chapter Model-based Linkage Analysis: All instances of lg should be replaced by log.
- Page 95, Chapter Model-based Linkage Analysis: In the formula after Eq. (6.4), the denominator in the LOD score function should be $0.5^{12}$ in the first expression and $0.5^{11}$ in the second expression.
- Page 96, Chapter Model-based Linkage Analysis: Section 6.1.3.2 has to be replaced by the following section:

If the mother is homozygous at one or both markers, one cannot determine whether a maternal recombination occurs. Consequently, the mother is not informative for linkage for haplotype combinations $H_{1}, H_{2}$, and $H_{3}$. Therefore, the situation is reduced to the setting considered before. The mother is heterozygous at both loci for haplotype combinations $H_{4}$ and $H_{5}$, and thus informative for linkage. For $H_{1}$ there are $n=12$ informative paternal meioses, and the number of recombinations is $k=2$, as already determined in Section 6.1.1. For $H_{2}$ and $H_{3}$ there are only 7 informative paternal meioses. Specifically, for both $H_{2}$ and $H_{3}$ offspring being homozygous 1 at both marker loci do not exhibit a recombination, while those being heterozygous at both marker loci are not informative for linkage. The offspring which is heterozygous at the first locus and homozygous at the second is informative for $\mathrm{H}_{2}$ and shows a recombination, but it is not informative for $H_{3}$. In contrast, the offspring that is homozygous at the first but heterozygous at the second locus is recombinant and informative for $H_{3}$ but not informative for $H_{2}$.
It is slightly more complicated to count the number of recombinants and nonrecombinants for $H_{4}$ and $H_{5}$. We therefore display the relevant parts of the corresponding pedigrees in Figure 6.4. Six offspring are homozygous for the 1 allele at both loci. Their haplotypes therefore are ${ }_{1}^{1}$. For $H_{4}$, there are twelve informative meioses for these six offspring showing a total of twelve non-recombinants, while they have a total of six recombinants and six non-recombinants for $H_{5}$. Four offspring are heterozygous at both loci. Because for $H_{4}$ and $H_{5}$ both parents are also heterozygous at both marker loci, the phase cannot be determined in the offspring. Both phases are equally likely if we assume linkage equilibrium between loci. For $H_{4}$ each of these offspring have either two non-recombinants or two recombinants, and exactly one recombinant and one non-recombinant for $H_{5}$.
Finally, the two remaining offspring who are heterozygous 12 at exactly one locus both have one recombinant and one non-recombinant for $H_{4}$. For $H_{5}$ they are either recombinant or non-recombinant for both meioses.
The kernels of the likelihoods corresponding to $H_{1}$ and $H_{5}$ thus are:
$L_{1}(\theta)=\theta^{2}(1-\theta)^{10}$

$$
\begin{equation*}
L_{2}(\theta)=\underbrace{(1-\theta)^{6}}_{6 \times \text { hom at both }} \cdot \underbrace{\theta}_{1 \times \text { het at first }} \tag{1}
\end{equation*}
$$

$$
L_{3}(\theta)=\underbrace{(1-\theta)^{6}}_{6 \times \text { hom at both }} \cdot \underbrace{\theta}_{1 \times \text { het at second }}
$$

$$
L_{4}(\theta)=\underbrace{\left((1-\theta)^{2}\right)^{6}}_{6 \times \text { hom at both }} \cdot \underbrace{\left(\theta^{2}+(1-\theta)^{2}\right)^{4}}_{4 \times \text { het at both }} \cdot \underbrace{(\theta(1-\theta))^{2}}_{2 \times \text { het at one }}
$$

$$
L_{5}(\theta)=\underbrace{(\theta(1-\theta))^{6}}_{6 \times \text { hom at both }} \cdot \underbrace{(\theta(1-\theta))^{4}}_{4 \times \text { het at both }} \cdot \underbrace{\left(\theta^{2}+(1-\theta)^{2}\right)^{2}}_{2 \times \text { het at one }}
$$

- Page 98, Chapter Model-based Linkage Analysis: After the last line in section 6.1.3.3 (line 5) the following text passage has to be inserted:

Finally, we want to stress that we do not weight haplotype frequencies in Eq. (6.6) with respect to the observed number of alleles in the offspring. Alternatively, one could use the Bayes formula for updating maternal haplotype probabilities given the observed number of 1-alleles at the two loci in the offspring.

In fact, this latter approach is followed in some software packages.

- Page 98, Chapter Model-based Linkage Analysis: Line 17-19, in the middle of Example 6.2, it should be

$$
\begin{aligned}
L(\theta) \approx & 0.0175 \cdot \theta^{2}(1-\theta)^{10} \\
& +0.3509 \cdot \theta(1-\theta)^{6} \\
& +0.3158 \cdot \theta^{2}(1-\theta)^{14}\left(\theta^{2}+(1-\theta)^{2}\right)^{4} \\
& +0.3158 \cdot \theta^{10}(1-\theta)^{10}\left(\theta^{2}+(1-\theta)^{2}\right)^{2}
\end{aligned}
$$

- Page 98, Chapter Model-based Linkage Analysis: In the last line of Example 6.2, it should be $\theta=0.14$.
- Page 98, Chapter Model-based Linkage Analysis: Table 6.1 has to be replaced by

| $\theta$ | 0.0 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $L O D$ score | $-\infty$ | 0.08 | 0.68 | 0.84 | 0.83 | 0.66 | 0.38 |

- Page 99, Chapter Model-based Linkage Analysis: Fig. 6.5 has to be replaced by Figure 1.


Figure 1: Kernel of the likelihood function from Example 6.2. Five haplotype combinations are possible in the mother (see Section 6.1.3.1), leading to kernels $L_{1}$ to $L_{5}$. Likelihood kernel $L_{1}$ has a weight of $0.0175 . L_{2}$ and $L_{3}$ are identical and, together, have a weight of 0.3509. $L_{4}$ and $L_{5}$ both have a weight of 0.3158 . The joint kernel of the likelihood function is the weighted average of $L_{1}$ to $L_{5}$.

- Page 100, Chapter Model-based Linkage Analysis: Fig. 6.6 has to be replaced by Figure 2.
- Page 102, Chapter Model-based Linkage Analysis: In line 15 , it should be $f_{0}=0, f_{1}=f_{2}=1$.
- Page 103, Chapter Model-based Linkage Analysis: In line 3 of the legend to Figure 6.9, "a" should be deleted.
- Page 103, Chapter Model-based Linkage Analysis: In line 11, it should be $f_{0}=f_{1}=0, f_{2}=1$.
- Page 106, Chapter Model-based Linkage Analysis: Last line, replace "most right" and "most left" with "rightmost" and "leftmost".
- Page 107, Chapter Model-based Linkage Analysis: Line 15 should be "If markers are spread out wide with inter-marker distances of $5-10 c M$, the marker".


Figure 2: Kernel of the likelihood function from Example 6.3. Five haplotype combinations are possible in the mother (see Section 6.1.3.1), leading to kernels $L_{1}$ to $L_{5}$. In the left side of the figure, $p_{1}=p_{2}=0.1$ is assumed, and the joint likelihood is dominated by $L_{4}$. In the right side of the figure, haplotype-specific kernels and the joint kernel of the likelihood are shown for $p_{1}=p_{2}=0.9$. It can be seen that the kernel is mostly influenced by likelihood kernels $L_{2}$ and $L_{3}$, which are all identical.

- Page 107, Chapter Model-based Linkage Analysis: In line 5 from bottom make "disease" plural to write "For complex genetic diseases ...".
- Page 109, Chapter Model-based Linkage Analysis: Replace each "if" with "for".
- Page 110, Chapter Model-based Linkage Analysis: Section 6.2.3, paragraph 3, line 2: "parental" should be "paternal".
- Page 110, Chapter Model-based Linkage Analysis: Offspring 2 in the right pedigree in Figure 6.12 should have the alleles 1 and 4 instead of 1 and 3.
- Page 112, Chapter Model-based Linkage Analysis: In line 7 in paragraph 2, "meiosis" should be exchanged by "meioses".
- Page 124, Chapter Model-free Linkage Analysis: In line 19, the word dominant has to be removed without substitution.
- Page 124, Chapter Model-free Linkage Analysis: In line 22, the term $F D$ is the abbreviation for family data that is available. Furthermore, $\Psi=\theta^{2}+(1-\theta)^{2}$. $\Psi$ points to the fact that either two or no recombination occurs per parent in the offspring.
- Page 126, Chapter Model-free Linkage Analysis: Line 3 of Example 7.1 begins with "the right part of the table."
- Page 127, Chapter Model-free Linkage Analysis: The section after Example 7.1 needs to be replaced by the following:
In almost all applications, the IBD value cannot be determined unambiguously for all sib-pairs. Therefore, the MLS statistic needs an extension that allows for incompletely informative ASPs. In this case, the likelihood is the probability for the genotypes of a sib-pair given that the sib-pair is an ASP. Using the law of total probability, the likelihood of sib-pair $i$ is $L_{i}=\sum_{j=0}^{2} w_{i j} z_{j}$, where $z_{j}$ is the

Table 1: Sample data for illustrating the use of the affected sib-pair (ASP) statistics. $\hat{z}_{i j}$ denotes the estimated probability that ASP $i$ shares $j$ alleles identical by descent (IBD), and $\hat{w}_{i j}$ denotes the weight of allele $j$ from ASP $i$ to the maximum LOD score test statistic. We assume that both parents have been genotyped. The hash symbol \# denotes number.

| \# ASPs | $\hat{z}_{i 0}$ | $\hat{z}_{i 1}$ | $\hat{z}_{i 2}$ | $\hat{w}_{i 0}$ | $\hat{w}_{i 1}$ | $\hat{w}_{i 2}$ | \# ASPs | $\hat{z}_{i 0}$ | $\hat{z}_{i 1}$ | $\hat{z}_{i 2}$ | $\hat{w}_{i 0}$ | $\hat{w}_{i 1}$ | $\hat{w}_{i 2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 21 | 0 | 0 | 1 | 0 | 0 | 1 | 25 | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{2}{3}$ | $\frac{1}{3}$ |
| 13 | 0 | 1 | 0 | 0 | 1 | 0 | 5 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| 4 | 1 | 0 | 0 | 1 | 0 | 0 | 7 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{1}{3}$ | $\frac{2}{3}$ | 0 |
|  |  |  |  |  |  |  | 6 | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |

unknown probability of an ASP sharing $j$ alleles IBD, and $w_{i j}$ is the probability for the marker genotypes, say $g_{\text {off }, i}$ of sib-pair $i$ given that they share $j$ alleles IBD. The joint likelihood is the product of the likelihood over all sib-pairs,
$L=\prod_{i=1}^{n}\left(w_{i 0} \cdot z_{0}+w_{i 1} \cdot z_{1}+w_{i 2} \cdot z_{2}\right)$.
The MLS subsequently is
MLS $=\lg \frac{\prod_{i=1}^{n}\left(\hat{w}_{i 0} \cdot \hat{z}_{0}+\hat{w}_{i 1} \cdot \hat{z}_{1}+\hat{w}_{i 2} \cdot \hat{z}_{2}\right)}{\prod_{i=1}^{n}\left(\hat{w}_{i 0} \cdot \frac{1}{4}+\hat{w}_{i 1} \cdot \frac{1}{2}+\hat{w}_{i 2} \cdot \frac{1}{4}\right)}$,
which is asymptotically $\chi_{2}^{2}$ distributed.
The weights $w_{i j}$ are constant if parental genotypes are available. For example, consider a nuclear family $i$ where all subjects, i.e., both parents and both offspring, are heterozygous 12 . Here, $\hat{z}_{i 2}=P\left(\mathrm{IBD}_{i}=2 \mid g_{\text {off }, i}\right)=\hat{z}_{i 0}=P\left(\operatorname{IBD}_{i}=\right.$ $\left.0 \mid g_{\mathrm{off}, i}\right)=\frac{1}{2}$. Application of the Bayes formula yields
$w_{i j}=P\left(g_{\text {off }, i} \mid \mathrm{IBD}_{i}=j\right)=\frac{P\left(\mathrm{IBD}_{i}=j \mid g_{\text {off }, i}\right) P\left(g_{\text {off }, i}\right)}{P\left(\mathrm{IBD}_{i}=j\right)}$.
In our example family, the first term of the numerator on the right hand side of Eq. (3) is estimated as $\frac{1}{2}$, thus identical for $\hat{w}_{i 0}$ and $\hat{w}_{i 2}$. The second term is independent of the IBD status, thus also equal. Finally, the denominator is $P\left(\operatorname{IBD}_{i}=0\right)=P\left(\operatorname{IBD}_{i}=2\right)=\frac{1}{4}$ for both $\hat{w}_{i 0}$ and $\hat{w}_{i 2}$. The estimated weights $\hat{w}_{i 0}$ and $\hat{w}_{i 2}$ are therefore identical for both IBD values and equal to $\frac{1}{2}$. Finally, $\hat{w}_{i 1}=0$ because $\hat{z}_{i 1}=0$. In this example, the estimated weights $w_{i j}$ indeed represent a probability distribution. We stress, however, that only the relative values of the $w_{i j}$ for $j=0,1,2$ are important, not their absolute values.
Consider a different example and assume that one parent is homozygous 11 , that the other parent is heterozygous 23 , and that both offspring are heterozygous 1 3. In this case, $\hat{z}_{i 1}=\hat{z}_{i 2}=\frac{1}{2}$ but $\hat{w}_{i 1}=2 \hat{w}_{i 2}$ because the denominator of Eq. (3) is $\frac{1}{2}$ for $\mathrm{IBD}_{i}=2$ and $\frac{1}{4}$ for $\mathrm{IBD}_{i}=1$.

Finally, we note that for non-informative ASPs, where both parents are homozygous at the marker locus, the estimated weights are $\hat{w}_{i j}=\frac{1}{3}$ for all $j=0,1,2$. Note that these non-informative contribute to both the MLS statistic of Eq. (2) and to the estimated IBD probabilities in the entire sample.
These sample estimates $\hat{z}_{0}, \hat{z}_{1}$, and $\hat{z}_{2}$ can be determined with an expectation maximization (EM) algorithm, and the update formula from step $k$ to $k+1$ is $\hat{z}_{j}^{(k+1)}=\frac{1}{n} \sum_{i=1}^{n} \frac{\hat{w}_{i j} \hat{z}_{j}^{(k)}}{\hat{w}_{i 0} \hat{z}_{0}^{(k)}+\hat{w}_{i 1} \hat{z}_{1}^{(k)}+\hat{w}_{i 2} \hat{z}_{2}^{(k)}}$.
Its application is illustrated in Example 7.2. Furthermore, we illustrate the
computation of the MLS using all example data from Table 1 in Example 7.3.

Table 2: Estimated probabilities $\hat{w}_{i j}$ for marker genotypes given identical by descent status for ten affected sib-pairs. $\hat{w}_{i j}$ are used for estimating identical by descent probabilities in the sample.

| $i$ | $\hat{w}_{i 0}$ | $\hat{w}_{i 1}$ | $\hat{w}_{i 2}$ | $i$ | $\hat{w}_{i 0}$ | $\hat{w}_{i 1}$ | $\hat{w}_{i 2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 0 | 0 | 1 | 6 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 | 7 | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| 3 | 0 | 1 | 0 | 8 | 0 | $\frac{2}{3}$ | $\frac{1}{3}$ |
| 4 | 0 | 1 | 0 | 9 | $\frac{1}{3}$ | $\frac{2}{3}$ | 0 |
| 5 | 0 | 1 | 0 | 10 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |

- Page 127, Chapter Model-free Linkage Analysis: Example 7.2 needs to be replaced by the following:
We consider $n=10$ ASPs with estimated weights $\hat{w}_{i j}$ as given in Table 2. As starting values we choose $\hat{z}_{0}^{(0)}=0.25, \hat{z}_{1}^{(0)}=0.5$, and $\hat{z}_{2}^{(0)}=0.25$. The first update gives $\hat{z}_{0}^{(1)}=0.1950, \hat{z}_{1}^{(1)}=0.5100$, and $\hat{z}_{2}^{(1)}=0.2950$. Using the update formula, we proceed in estimating $\hat{z}_{j}^{(2)}$, which turns out to be $\hat{z}_{0}^{(2)} \approx 0.1763$, $\hat{z}_{1}^{(2)} \approx 0.5125$, and

$$
\begin{aligned}
\hat{z}_{2}^{(2)}= & \frac{1}{10} \sum_{i=1}^{10} \frac{\hat{w}_{i 2} \cdot \hat{z}_{2}^{(1)}}{\hat{w}_{i 0} \hat{z}_{0}^{(1)}+\hat{w}_{i 1} \hat{z}_{1}^{(1)}+\hat{w}_{i 2} \hat{z}_{2}^{(1)}}=\frac{1}{10} \sum_{i=1}^{10} \frac{\hat{w}_{i 2} \cdot 0.1950}{\hat{w}_{i 0} 0.1950+\hat{w}_{i 1} 0.5100+\hat{w}_{i 2} 0.2950} \\
= & \frac{1}{10} \cdot\left(\frac{1 \cdot 0.2950}{0 \cdot 0.1950+0 \cdot 0.5100+1 \cdot 0.2950}+\frac{1 \cdot 0.2950}{0 \cdot 0.1950+0 \cdot 0.5100+1 \cdot 0.2950}\right. \\
& +\frac{0 \cdot 0.2950}{0 \cdot 0.1950+1 \cdot 0.5100+0 \cdot 0.2950}+\frac{0 \cdot 0.2950}{0 \cdot 0.1950+1 \cdot 0.5100+0 \cdot 0.2950} \\
& +\frac{0 \cdot 0.250}{0 \cdot 0.1950+1 \cdot 0.5100+0 \cdot 0.2950}+\frac{0 \cdot 0.2950}{1 \cdot 0.1950+0 \cdot 0.5100+0 \cdot 0.2950} \\
& +\frac{\frac{1}{3} \cdot 0.2950}{\frac{1}{3} \cdot 0.1950+\frac{1}{3} \cdot 0.5100+\frac{1}{3} \cdot 0.2950}+\frac{\frac{1}{3} \cdot 0.2950}{0 \cdot 0.1950+\frac{2}{3} \cdot 0.5100+\frac{1}{3} \cdot 0.2950} \\
& \left.+\frac{0.0 .2950}{\frac{1}{3} \cdot 0.1950+\frac{2}{3} \cdot 0.5100+0 \cdot 0.2950}+\frac{\frac{1}{2} \cdot 0.2950}{\frac{1}{2} \cdot 0.1950+0 \cdot 0.5100+\frac{1}{2} \cdot 0.2950}\right) \\
\approx & 0.3121 .
\end{aligned}
$$

Six updates are required to estimate the IBD probabilities with a precision of $10^{-4}$. The update steps are given in Table 3. The final ML IBD estimates are $\hat{z}_{0} \approx 0.1640, \hat{z}_{1} \approx 0.5137$, and $\hat{z}_{2} \approx 0.3222$.

Table 3: Updates from the expectation maximization algorithm for estimating the identical by descent (IBD) probabilities in the sample of ten affected sib-pairs given in Table 2.

|  | IBD probability estimates |  |  |
| :--- | :---: | :---: | :---: |
| Update step | $\hat{z}_{0}$ | $\hat{z}_{1}$ | $\hat{z}_{2}$ |
| 0 | (Initialization) | 0.2500 | 0.5000 |
| 1 | 0.1950 | 0.5100 | 0.2500 |
| 2 | 0.1753 | 0.5125 | 0.2950 |
| 3 | 0.1681 | 0.5133 | 0.3121 |
| 4 | 0.1654 | 0.5135 | 0.3185 |
| 5 | 0.1644 | 0.5136 | 0.3210 |
| 6 | 0.1640 | 0.5137 | 0.3219 |

- Page 129, Chapter Model-free Linkage Analysis: Example 7.3 needs to be replaced by the following:
In this example, we illustrate the calculation of the MLS using all data from Table 1. In a first step, the EM update formula (4) is used for estimating $\hat{z}_{j}$. The parameter estimates for the total sample of 81 ASPs are $\hat{z}_{0}=0.0684, \hat{z}_{1}=$ 0.4907 , and $\hat{z}_{2}=0.4409$. The MLS for these data is

$$
\begin{aligned}
\text { MLS }=\lg & \underbrace{\frac{(0 \cdot 0.0684+0 \cdot 0.4907+1 \cdot 0.4409)^{21}}{\left(0 \cdot \frac{1}{4}+0 \cdot \frac{1}{2}+1 \cdot \frac{1}{4}\right)^{21}}}_{21 \text { families with IBD }=2} \cdot \underbrace{\frac{(0 \cdot 0.0684+1 \cdot 0.4907+0 \cdot 0.4409)^{13}}{\left(0 \cdot \frac{1}{4}+1 \cdot \frac{1}{2}+0 \cdot \frac{1}{4}\right)^{13}}}_{13 \text { families with IBD }=1} \\
& \cdot \underbrace{\frac{(1 \cdot 0.0684+0 \cdot 0.4907+0 \cdot 0.4409)^{4}}{\left(1 \cdot \frac{1}{4}+0 \cdot \frac{1}{2}+0 \cdot \frac{1}{4}\right)^{4}}}_{4 \text { families with IBD }=0} \cdot \underbrace{\frac{\left(0 \cdot 0.0684+\frac{2}{3} \cdot 0.4907+\frac{1}{3} \cdot 0.4409\right)^{25}}{\left(0 \cdot \frac{1}{4}+\frac{2}{3} \cdot \frac{1}{2}+\frac{1}{3} \cdot \frac{1}{4}\right)^{25}}}_{25 \text { families with IBD }=1 \text { or } 2} \\
& \cdot \underbrace{\frac{\left(\frac{1}{3} \cdot 0.0684+\frac{2}{3} \cdot 0.4907+0 \cdot 0.4409\right)^{7}}{\left(\frac{1}{3} \cdot \frac{1}{4}+\frac{2}{3} \cdot \frac{1}{2}+0 \cdot \frac{1}{4}\right)^{7}}} \cdot \underbrace{\frac{\left(\frac{1}{2} \cdot 0.0684+0 \cdot 0.4907+\frac{1}{2} \cdot 0.4409\right)^{5}}{\left(\frac{1}{2} \cdot \frac{1}{4}+0 \cdot \frac{1}{2}+\frac{1}{2} \cdot \frac{1}{4}\right)^{5}}}
\end{aligned}
$$

7 families with $\mathrm{IBD}=0$ or 15 families with $\mathrm{IBD}=0$ or 2

$$
\cdot \underbrace{\left.\frac{\left(\frac{1}{3} \cdot 0.0684+\frac{1}{3} \cdot 0.4907+\frac{1}{3} \cdot 0.4409\right)^{6}}{\left(\frac{1}{3} \cdot \frac{1}{4}+\frac{1}{3} \cdot \frac{1}{2}+\frac{1}{3} \cdot \frac{1}{4}\right)^{6}}\right)}
$$

6 non informative families
$\approx 3.7283$.
Thus, the $L O D$ score decreased from 3.86 to 3.73 in this example by adding the partly informative families compared with the use of the completely informative families from Example 7.1. The $L O D$ score of 3.73 corresponds to $T \approx 17.1694$, giving $p=1.87 \cdot 10^{-4}$ by using the $\chi_{2}^{2}$ distribution. The question is why the $L O D$ score dropped although families were added to the entire sample. This can be easily explained by looking at the IBD estimates. While the estimated probability for an ASP to share 2 alleles IBD was approximately $55 \%$ in the completely informative families, this fraction decreased to $44 \%$ in the entire sample. We finally note that the completely non-informative families are omitted from calculations in some of the freely available software packages.

- Page 131, Chapter Model-free Linkage Analysis: The first sentence in paragraph 2 of Section 7.3.1.4 should read: "We illustrate the use of the MLS with two simple examples."
- Page 141, Chapter Model-free Linkage Analysis: Table 7.7, in footnote (a) the 90th percentile of the distribution of the BMI needs to be referred to as the $10 \%$ most obese.
- Page 142, Chapter Model-free Linkage Analysis: In section 7.5.2, formula 7.12 should read: $\lambda_{R}-1=\frac{1}{K^{2}}\left(2 \cdot c_{R} \sigma_{a}^{2}+z_{R, 2} \sigma_{d}^{2}\right)$
- Page 163, Chapter Model Free Linkage Analysis for Quantitative Traits: In line 13 (formula 8.6), it should be

$$
\mathbb{E}\left(y_{i} \mid \mathrm{IBD}_{t}\right)=\sigma_{\epsilon}^{2}+2 \sigma_{g}^{2}-2 \sigma_{a}^{2} \tau_{t, i}-2 \sigma_{d}^{2} z_{t, i 2}=\alpha+\beta^{*} \tau_{t, i}+\gamma^{*} z_{t, i 2}
$$

- Page 186, Chapter Model Free Linkage Analysis for Quantitative Traits, Problem 8.3: The genotype of the mother of family 9) should be 33 rather than 12. At the same time, both offspring genotypes should be 13 .
- Pages 192-193, Chapter Fundamental Concepts of Association Analysis: Ex-
ample 9.1 needs to be replaced with the following:
In the association study by Reich and colleagues [360], 231 patients suffering from psoriasis were compared with 345 healthy controls. All probands were genotyped on a number of SNPs in the genes encoding for tumor necrosis factor$\alpha$ (TNFA). For two SNPs, TNFA-238 and TNFA-308, the genotypes of the healthy controls' typings are displayed in Table 9.4. Note that at TNFA-238, no proband was homozygous for the $A$ allele.
To establish the LD between the two SNPs, the allele frequencies $p_{G, 238}$ at TNFA-238 and $p_{G, 308}$ at TNFA-308, which are the allele frequencies of the respective $G$ allele, respectively, as well as the haplotype frequencies need to be estimated. Given Hardy-Weinberg equilibrium, the allele frequencies can be calculated easily from the given genotype frequencies and are given by

$$
\begin{array}{cl}
\hat{p}_{G, 238}=\frac{2 \cdot 316+1 \cdot 29}{2 \cdot 345} \approx 0.957971, \quad \hat{q}_{G, 238}=\frac{2 \cdot 0+1 \cdot 29}{2 \cdot 345} \approx 0.042029 \\
\hat{p}_{G, 308}=\frac{2 \cdot 238+1 \cdot 103}{2 \cdot 345} \approx 0.839130, & \hat{q}_{G, 308}=\frac{2 \cdot 4+1 \cdot 103}{2 \cdot 345} \approx 0.160870
\end{array}
$$

As pointed out, the determination of the haplotype frequencies is more difficult because only genotype frequencies in the sample had been ascertained. To carry on with this example, the expectation maximization (EM) algorithm described in Chapter 12 was employed to estimate the haplotype frequencies, and the results are shown in Table 9.5. Given these frequencies, different LD statistics can now be calculated, for example,
$\mathcal{D}_{G G} \approx 0.797102-0.957971 \cdot 0.839130 \approx-0.0067602$, and
$\mathcal{D}_{G G}^{\prime} \approx \frac{-0.0067602}{-0.160870 \cdot 0.042029} \approx 0.999851$.
It should be noted that the occurrence of the haplotype $G G$ is rarer than expected from the marginal distributions, so that $\mathcal{D}_{G G}$ becomes negative. However, given the marginal frequencies, the maximum of $\mathcal{D}_{G G}$ is $(-0.160870$. 0.042029), so that $\mathcal{D}_{G G}^{\prime}$ becomes positive.

- Pages 192-193, Chapter Fundamental Concepts of Association Analysis: In line with the previous point, Tables 9.4 and 9.5 are replaced by:

|  |  | TNFA-308 |  |  |  |
| :--- | :--- | ---: | ---: | ---: | ---: |
|  |  | $G G$ | $G A$ | $A A$ | Total |
| TNFA-238 | $G G$ | 213 | 99 | 4 | 316 |
|  | $G A$ | 25 | 4 | 0 | 29 |
|  | Total | 238 | 103 | 4 | 345 |


| TNFA-308 |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| $G$ | $A$ |  |  |  |
|  | $G$ | 0.797102 | 0.160870 | 0.957972 |
|  | $A$ | 0.042029 | $<0.00001$ | 0.042029 |
|  |  | 0.839131 | 0.160870 | 1 |

- Page 201, Chapter Association Analysis in Unrelated Individuals: The sentence preceding Eq. (10.1) should be replaced by:
Instead of using the OR as a quotient, however, the squared standardized difference has more appealing distribution properties. Hence, one such standard $\chi^{2}$ test has the following form:
- Page 212, Chapter Association Analysis With Unrelated Individuals: In line 21 and 22 , the value " 0.4549 " in the formula for the estimate of inflation factor should be 0.456 .
- Page 235, Chapter Family-based Association Analysis: Fig. 11.5 has to be replaced by Figure 3.


Figure 3: Illustrative data for the computation of the reconstruction combined transmission disequilibrium test (RC-TDT). Eight families of four different types as described in Table 11.12 have been genotyped at one diallelic marker with the alleles $A$ and $a$.

- Page 251, Chapter Haplotypes: In line 4, reference to Schaid [377] is incorrect. Instead, reference should be to Schaid DJ (2004) Evaluating associations of haplotypes with traits. Genetic Epidemiology 27: 348-364.
- Page 279, Chapter Solutions to Study Problems: In line 18, Solution 2.1.4, it should be "X-chromosomal recessive".
- Page 282, Chapter Solutions to Study Problems: In the beginning of paragraph 2 of Solution 4.5, $\epsilon=0.01$ should be $\epsilon=0.1$.
- Page 282, Chapter Solutions to Study Problems: In the middle part of the first equation in paragraph 2 of Solution 4.5 , it should be $0.4400^{2}$ instead of 0.4400 . The second to last line of the calculations should be replaced with: $\left(\left(0.4400+0.3467^{2}\right)^{2}-4 \cdot 0.3467\right)$
- Page 285, Chapter Solutions to Study Problems: Solution 5.6 should be deleted.
- Page 286, Chapter Solutions to Study Problems: In line 2, it should be $L_{1}(\theta)=\underbrace{(1-\theta)^{12}}_{O_{1}} \cdot \underbrace{\left(0.5 \theta^{2}+0.5(1-\theta)^{2}\right)^{4}}_{O_{2}} \cdot \underbrace{\theta(1-\theta)}_{O_{3}} \cdot \underbrace{\theta(1-\theta)}_{O_{4}}$.
- Page 286, Chapter Solutions to Study Problems: In line 10, it should be $L_{2}(\theta)=\underbrace{\theta^{6}(1-\theta)^{6}}_{O_{1}} \cdot \underbrace{\theta^{4}(1-\theta)^{4}}_{O_{2}} \cdot \underbrace{\left(0.5 \theta^{2}+0.5(1-\theta)^{2}\right)}_{O_{3}} \cdot \underbrace{\left(0.5 \theta^{2}+0.5(1-\theta)^{2}\right)}_{O_{4}}$.
- Page 289, Chapter Solutions to Study Problems: Solution 7.1 reads: "In the first pedigree IBS and IBD values are both 1. In the second pedigree they are 2."
- Page 289, Chapter Solutions to Study Problems: The update formula for Solution 7.2.1 reads:

$$
\begin{aligned}
& \hat{z}_{0}^{(k+1)}=\frac{1}{81}\left(4+7 \frac{\frac{1}{3} \hat{z}_{0}^{(k)}}{\frac{1}{3} \hat{z}_{0}^{(k)}+\frac{2}{3} \hat{z}_{1}^{(k)}}+5 \frac{\frac{1}{2} \hat{z}_{0}^{(k)}}{\frac{1}{2} \hat{z}_{0}^{(k)}+\frac{1}{2} \hat{z}_{2}^{(k)}}+6 \hat{z}_{0}^{(k)}\right) \\
& \hat{z}_{1}^{(k+1)}=\frac{1}{81}\left(13+25 \frac{\frac{2}{3} \hat{z}_{1}^{(k)}}{\frac{2}{3} \hat{z}_{1}^{(k)}+\frac{1}{3} \hat{z}_{2}^{(k)}}+7 \frac{\frac{2}{3} \hat{z}_{1}^{(k)}}{\frac{1}{3} \hat{z}_{0}^{(k)}+\frac{2}{3} \hat{z}_{1}^{(k)}}+6 \hat{z}_{1}^{(k)}\right) \\
& \hat{z}_{2}^{(k+1)}=\frac{1}{81}\left(21+25 \frac{\frac{1}{3} \hat{z}_{2}^{(k)}}{\frac{2}{3} \hat{z}_{1}^{(k)}+\frac{1}{3} \hat{z}_{2}^{(k)}}+5 \frac{\frac{1}{2} \hat{z}_{2}^{(k)}}{\frac{1}{2} \hat{z}_{0}^{(k)}+\frac{1}{2} \hat{z}_{2}^{(k)}}+6 \hat{z}_{2}^{(k)}\right) .
\end{aligned}
$$

- Page 290, Chapter Solutions to Study Problems: Table S. 7 needs to be replaced by the following Table 4.

Table 4: Updates from the expectation maximization algorithm for estimating the identical by descent (IBD) probabilities in the sample of the 81 affected sib-pairs given in Table 1.

| Update <br> step | IBD probability estimates |  | Update |  | IBD probability estimates |  |  |  |
| :--- | :---: | :---: | :---: | :--- | :---: | :---: | :---: | :---: |
|  | $\hat{z}_{0}$ | $\hat{z}_{1}$ | $\hat{z}_{2}$ | step |  | $\hat{z}_{0}$ | $\hat{z}_{1}$ | $\hat{z}_{2}$ |
| 0 | 0.10530 | 0.3421 | 0.5526 | 3 |  | 0.0689 | 0.4824 | 0.4485 |
| 1 | 0.07858 | 0.4314 | 0.4899 | 4 |  | 0.0684 | 0.4875 | 0.4439 |
| 2 | 0.07095 | 0.4685 | 0.4605 | 5 |  | 0.0683 | 0.4894 | 0.4421 |

- Page 290, Chapter Solutions to Study Problems: Solution 7.4.1 reads: "The estimate of the IBD distribution for the total sample prior to application of the EM algorithm gives $\hat{z}_{0}=0.0007, \hat{z}_{1}=0.2894$, and $\hat{z}_{2}=0.7099$. We can easily verify that this is a point in the possible triangle. First, $\hat{z}_{1}=0.2894$, thus $<0.5$. Second, $2 \hat{z}_{0}<\hat{z}_{1}$."
- Page 290, Chapter Solutions to Study Problems: In Table S.8, $\hat{z}_{i 2}$ for Family 7 is 0 .
- Page 293, Chapter Solutions to Study Problems: The estimate $\overline{\hat{\tau}}$ is 0.5 , not 0.45 . The correct value 0.5 is given in the last line of Table S.9.
- Page 299, Chapter Solutions to Study Problems: The CI(OR) in Solution 10.2 is [0.1408; 28.4159].
- Page 305 to 328 , Section References: References with exactly three authors erroneously include "et al." in the authors list.

