February 28, 2007

A Statistical Approach to Genetic Epidemiology - Corrections

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 68, Chapter Data Quality: In algorithm 4.5, step 2 (d) should read "If $Hom_O \geq Hom_P$, add 1 to C".
- 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 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 non-recombinants 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 $\frac{1}{1}$. 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 1 2 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:

$$
\begin{aligned}
L_1(\theta) &= \theta^2(1-\theta)^{10} \qquad (1)\\
L_2(\theta) &= \underbrace{(1-\theta)^6}_{6\times\text{hom at both}} \cdot \underbrace{\theta}_{1\times\text{het at first}}\\
L_3(\theta) &= \underbrace{(1-\theta)^6}_{6\times\text{hom at both}} \cdot \underbrace{\theta}_{1\times\text{het at second}}\\
L_4(\theta) &= \underbrace{\Big((1-\theta)^2\Big)^6}_{6\times\text{hom at both}} \cdot \underbrace{\Big(\theta^2+(1-\theta)^2\Big)^4}_{4\times\text{het at both}} \cdot \underbrace{\Big(\theta(1-\theta)\Big)^2}_{2\times\text{het at one}}\\
L_5(\theta) &= \underbrace{\Big(\theta(1-\theta)\Big)^6}_{6\times\text{hom at both}} \cdot \underbrace{\Big(\theta(1-\theta)\Big)^4}_{4\times\text{het at both}} \cdot \underbrace{\Big(\theta^2+(1-\theta)^2\Big)^2}_{2\times\text{het at one}}
\end{aligned}
$$

- 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

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

- 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

θ	0.0	0.01	0.05	0.1	0.2	0.3	0.4
LOD 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.
- 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 11, it should be

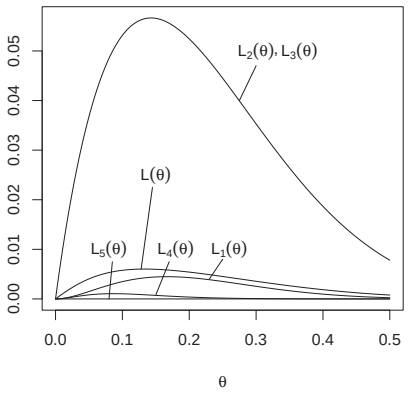

Figure 1: Kernel of the likelihood function from Example 6.2. Five haplotype combinations are possible in the mother (see Section 6.1.3.1), leading to kernels L_1 to L_5. Likelihood kernel L_1 has a weight of 0.0175. L_2 and L_3 are identical and, together, have a weight of 0.3509. L_4 and L_5 both have a weight of 0.3158. The joint kernel of the likelihood function is the weighted average of L_1 to L_5.

$f_0 = f_1 = 0, f_2 = 1$.

- 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 *FD* is the abbreviation for *family data*.
- Page 163, Chapter Model Free Linkage Analysis for Quantitative Traits: In line 13 (formula 8.6), it should be
$$\mathbb{E}(y_i|\text{IBD}_t) = \sigma^2_\epsilon + 2\sigma^2_g - 2\sigma^2_a\tau_{t,i} - 2\sigma^2_d z_{t,i2} = \alpha + \beta^*\tau_{t,i} + \gamma^* z_{t,i2}\,.$$
- Pages 192-193, Chapter Fundamental Concepts of Association Analysis: **Example 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-α (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
$$\hat{p}_{G,238} = \frac{2\cdot 316 + 1\cdot 29}{2\cdot 345} \approx 0.957971\,, \qquad \hat{q}_{G,238} = \frac{2\cdot 0 + 1\cdot 29}{2\cdot 345} \approx 0.042029\,,$$
$$\hat{p}_{G,308} = \frac{2\cdot 238 + 1\cdot 103}{2\cdot 345} \approx 0.839130\,, \qquad \hat{q}_{G,308} = \frac{2\cdot 4 + 1\cdot 103}{2\cdot 345} \approx 0.160870\,.$$
  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,
  $\mathcal{D}_{GG} \approx 0.797102 - 0.957971\cdot 0.839130 \approx -0.0067602$, and

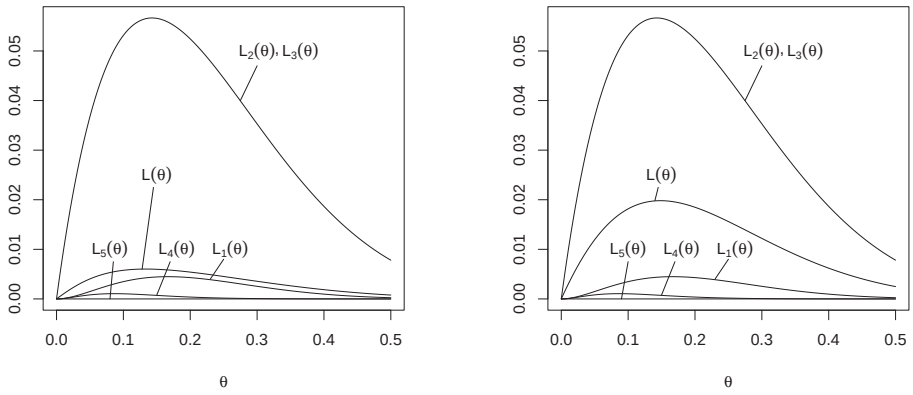

Figure 2: Kernel of the likelihood function from Example 6.3. Five haplotype combinations are possible in the mother (see Section 6.1.3.1), leading to kernels L_1 to L_5. In the left side of the figure, $p_1 = p_2 = 0.1$ is assumed, and the joint likelihood is dominated by L_4. In the right side of the 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.

$$\mathcal{D}'_{GG} \approx \frac{-0.0067602}{-0.160870 \cdot 0.042029} \approx 0.999851\,.$$

It should be noted that the occurrence of the haplotype GG is rarer than expected from the marginal distributions, so that $\mathcal{D}_{GG}$ becomes negative. However, given the marginal frequencies, the maximum of $\mathcal{D}_{GG}$ is $(-0.160870 \cdot 0.042029)$, so that $\mathcal{D}'_{GG}$ 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			
		GG	GA	AA	Total
TNFA-238	GG	213	99	4	316
	GA	25	4	0	29
	Total	238	103	4	345

		TNFA-308		
		G	A	
TNFA-238	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 χ^2 test has the following form:
- Page 208, Chapter Association Analysis in Unrelated Individuals: The last equation in the example should be replaced by:

$$\mathrm{OR}_A = 1 + \frac{0.07(2-1)}{0.3[(1-0.3) + (0.1(1-0.3) - 0.07)(2-1)]} = 1.33\,.$$

- 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.

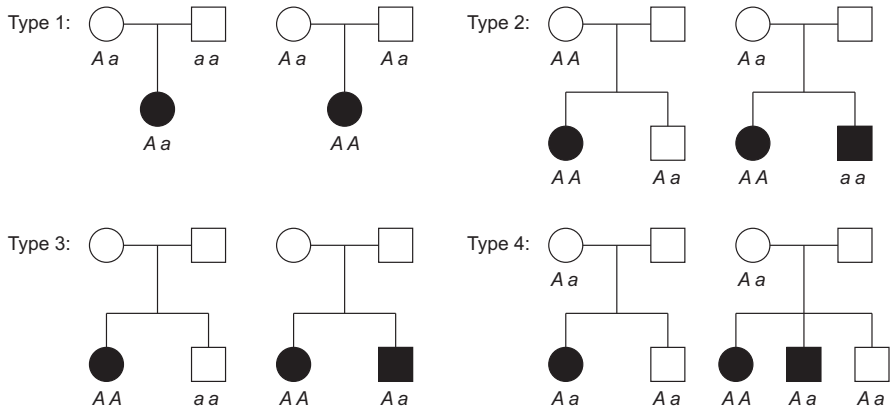

Figure 3: Illustrative data for the computation of the reconstruction combined transmission disequilibrium test (RC-TDT). Eight families of four different types as described in Table 11.12 have been genotyped at one diallelic marker with the alleles A and a.

- 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 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 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.