Linkage Newsletter

Vol. 1   No. 2   Dec. 1987

EDITORIAL

The response to the previous, first issue of the Newsletter was extremely positive. I received many support- ive letters from various countries -- thank you all for writing. This positive response reinforces my impression that the Newsletter fills a need in our scientific community.

Several readers of the last issue of the Newsletter inquired about availability of the computer programs men- tioned. While I will continue discussing software developed by anyone, I can only distribute programs developed by me or my collaborators (a few exceptions are some generally avail- able programs of a more technical nature, such as the Kermit program for computer communication, or the PC-WRITE text editor). Programs developed by other researchers will have to be ordered from them directly. Note that the Mapmaker program, mentioned in the last issue of the Newsletter, has just been described in detail (Genomics 1, 174-181, 1987).

Readers who would like to make a written contribution to the Newsletter are encouraged to write to me (I cannot, how- ever, absolutely guarantee publication). Texts longer than half a typewritten page are requested in the form of an ASCII (text) file on a floppy disk, 5 1/4", IBM PC format. The author's name will appear as given on the text submitted (initials or full name, with or without address).

When the International Workshop on Human Gene Mapping was held in Paris last summer, several useful computer pro- grams developed by human geneticists were demonstrated at the CEPH building. There were database programs for linkage analysis, programs to draw pedigrees on screen or on a matrix printer, and so on. I would like to compile and maintain a list of available programs and publish this list at regular intervals as part of the Newsletter. To be able to do this, I need the cooperation of the authors of such programs -- if you have a computer program you would like to make available to other users please fill out the attached questionnaire and return it to me.

ANNOUNCEMENT

Dr. Sue Hodge submitted the following announcement regarding the QUAD computer program. It is reprinted here as sent:

We have recently identified an error in an earlier version of our computer program QUAD (Hodge et al., 1983, p. 1141). **All the quadratic interpolations themselves**, both one- and two-dimensional, are correct. However, the program calculated the standard errors and the variance-covariance matrix incorrectly. Specifically, the standard error given by the incorrect program is too large by a factor of $1.5174 \times \sqrt{\ln 10}$; that is, the standard error must be multiplied by 0.6590 to be correct. The error is easily corrected by changing two lines in the FORTRAN code. (Write us for details.)

To determine whether your version of QUAD contains the error, simply enter the following lod scores for one-dimensional interpolation:

θ	.01	.05	.10	.20	.30	.40	.50
lod	1.06	1.29	1.26	1.10	.94	.63	0

The correct standard error for this test case is .0555. If your program gives a value of .0842, then the correction needs to be made.

We have notified those individuals and centers who requested our program. However, we are publishing this notice to ensure that everyone using the program has been alerted.

Reference
Hodge SE, Anderson CE, Neiswanger K, Sparkes RS, Rimoin DL: The search for heterogeneity in insulin-dependent diabetes mellitus (IDDM): Linkage studies, two-locus models, and genetic heterogeneity. Am J Hum Genet 35:1139-1155, 1983.

SOFTWARE NOTES

Version 4.6 of the **Linkage** programs has been made available. It represents a major improvement over the previous version (3.5). Included in the package is a command shell program (LCP) which is menu-driven and builds a batch file that can then be run to carry out the calculations defined in the LCP program. Another program included in the package is a report generator. The whole package may be obtained on 18 floppy disks from Dr. Mark Lathrop in Salt Lake City.

For PC/AT users who are interested in analyzing general pedigrees only, the corresponding version 4.6 Linkage programs are available from me on 4 disks (3 without source code) (please write to me for ordering instructions). Unfortunate-

ly, I have so far been unable to compile the new program version on either an IBM mainframe (Pascal/VS) or on the DSI-32 coprocessor board.

Using a pedigree with 91 individuals (no loops) and 3 codominant markers I happened to work on, I ran the new and old versions of Ilink on different machines. For version 4.6, the running times include processing of the input file by the 'Unknown' program. The Ilink program took 7 iterations and a total of 39 function evaluations. On the IBM mainframe, it finished sooner; the time shown below is extrapolated to the number of iterations taken in the other runs.

Program version	Machine	Running time
3.5	IBM AT 8MHz	71 min. 30 sec.
3.5	IBM AT 8MHz using 80287	58 min. 58 sec.
3.5	IBM 4381-11 mainframe	4 min. 12 sec.
4.6	IBM AT 8MHz	9 min. 53 sec.
4.6	IBM AT 8MHz using 80287	9 min. 27 sec.

Mark Lathrop and Peter Cartwright in Salt Lake City are planning on sending me some notes on the new Linkage version for inclusion in the next issue of the Newsletter.

I modernized **PC-LIPED** somewhat and included various example data files but did not change the basic algorithms. The new features are the following:
- Lognormal and straight-line age-dependant penetrance are implemented as special locus types.
- PC-LIPED does most calculations in double precision.
- Output is printed to the screen and optionally to a file so that the program may be interrupted with Ctrl-C without loss of the results obtained up to the break.
- PC-LIPED now comes in 2 versions, one of them does and the other does not use the numeric coprocessor. Previously, the Microsoft library used in the compilation allowed detecting presence of the coprocessor so that it would automatically be used when present. However, sensing presence of the coprocessor apparently fails in some IBM PC clones so that on those machines, PC-LIPED did not run.
- The program is compiled with Microsoft Fortran v.4.01. However, execution speed is about the same as when compiled with version 3.13 of MS-Fortran.

Programs for linkage analysis on microcomputers should be used with a **numeric coprocessor** whenever possible. Both

Turbo Pascal and MS Fortran allow a much larger range of real numbers when the coprocessor is used than when it is not. Without it, with larger family sizes, the value of the likelihood can sometimes become too small resulting in an underflow. While LIPED (MS Fortran) reports the occurrence of an underflow as an error, the Linkage programs (Turbo Pascal) do not and simply set the likelihood (of one or more genotypes) equal to zero. In unfortunate cases, this can lead to the situation that a family appears to contain an inconsistency while in fact it does not, or that it falsely appears to be uninformative for linkage. The Linkage programs as distributed do not make use of the coprocessor which is fine for most applications, but as a safeguard against nondetectable underflows, people who have a coprocessor installed in their microcomputer are encouraged to recompile the programs using the Turbo Pascal compiler with 8087 support.

My collaborators have recently started using the DESQview program version 2.0 (Quarterdeck Office Systems, Santa Monica). Although we presently have only 640KB of memory available on the Compaq Deskpro 386, the program behaves beautifully. For example, while one is working with WordPerfect on a manuscript, DESQview allows running in the background a long Mlink job on the DSI-32 coprocessor board and at the same time uploading a file to the mainframe computer. DESQview is inexpensive, we can buy it here in New York for $79.

QUESTIONS AND ANSWERS

Q: The application of results from linkage analyses to risk calculation requires the use of confidence limits for the recombination fraction θ that was estimated. Due to the lack of knowledge about the true distribution of lod scores, $z(\theta)$, no confidence limits for θ can be found in a strictly statistical sense.

One of the approximations proposed to avoid this problem was the use of so-called 'one-unit-down' confidence regions comprising all θ's for which $z(\theta) > z(\theta_{max}) - 1$.

I have calculated the significance of this method for the simplest case where the observations are distributed as binomial $B(n,\theta)$ variables, that is, recombinants and non-recombinants can definitely be distinguished. I performed these calculations using the normal approximation for the standardized number, S_0, of recombinants:
$$S_0 = (S - n\theta)/\sqrt{[n\theta(1-\theta)]} \approx N(0,1),$$
where S = number of recombinants. The results showed that

for n between 1000 and 2000,

$$c = \min_{\theta < 0.5} P_\theta[z(\theta) > z(\theta max) - 1] \approx 85\%. \tag{1}$$

If 1.75 instead of 1 is chosen as the 'max-lod' difference then $c \approx 95\%$. I know that in the case of real linkage data the graph of lod scores vs. recombination fraction does not have peaks as sharp as in the case of $B(n,\theta)$. Therefore, the 'confidence regions' obtained by the 'one-unit-down' method might be much larger. Can you give any hints about how to deal with this problem? Do you know whether and by whom it was treated mathematically? (Michael Krawczak, Institute of Human Genetics, Gosslerstr. 12 d, D-3400 Göttingen, FRG).

A: There is a well-known close relationship between signi-ficance tests and confidence intervals in the sense that a confidence interval comprises all those parameter values, θ_0, that do not make the test of $H_0:\theta=\theta_0$ significant. In the present example, a confidence interval has been defined as consisting of all those θ-values for which $z(\theta) > z(\theta_{max}) - 1$. Denote its endpoints by θ_1 and θ_2, $\theta_1 < \theta < \theta_2$.

To avoid problems connected with approximating the bino-mial by the normal distribution, I calculated the values in the table on the next page using the binomial distribution, where k = number of recombinants, n = total number of recom-binants and nonrecombinants, α_1 and α_2 are the "error probab-ilities" to the left and right of the confidence interval, and $c = 1 - \alpha_1 - \alpha_2$ is the confidence coefficient. The table suggests that the "1-unit-down" method results in reliable confidence intervals, at least for phase-known double back-cross families. Analogous calculations for phase-unknown families with 2 offspring each confirm the general findings of the table. The only instance possibly leading to c<0.90 appears to occur with of a maximum lod score smaller than 1 and the restriction, $\theta_2 = 0.5$, but that case is specifically excluded from consideration.

Strictly speaking, the 'one-unit-down' method leads to a so-called support interval as it is based entirely on the log likelihood (= support) and not on any test theory. The notion of support intervals has been advanced by A.W.F. Edwards, "Likelihood", New York: Cambridge University Press, 1972. However, such support intervals may also be inter-preted as confidence intervals (see above). The values for α_1 and α_2, above, were calculated based on J. Pfanzagl: "Allgemeine Methodenlehre der Statistik II", Sammlung Gö-schen, Walter de Gruyter, Berlin 1966, p.114. A more demand-ing treatment of confidence intervals may be found in statis-tics textbooks such as C.R. Rao: "Linear statistical infer-ence and its applications", Wiley, New York. (JO)

k	n	θ_1	θ_2	α_1	α_2	c
0	10	0.000	0.206	0.000	0.100	0.900
1	10	0.004	0.403	0.040	0.044	0.915
0	20	0.000	0.109	0.000	0.100	0.900
1	20	0.002	0.222	0.039	0.044	0.916
2	20	0.014	0.298	0.031	0.036	0.933
5	20	0.088	0.484	0.026	0.029	0.945
0	50	0.000	0.045	0.000	0.100	0.900
1	50	0.001	0.094	0.039	0.044	0.917
2	50	0.006	0.128	0.031	0.036	0.933
5	50	0.033	0.214	0.025	0.029	0.947
10	50	0.099	0.337	0.022	0.025	0.953
0	100	0.000	0.023	0.000	0.100	0.900
1	100	0.000	0.048	0.039	0.044	0.917
2	100	0.003	0.066	0.030	0.036	0.934
5	100	0.016	0.111	0.024	0.029	0.947
10	100	0.048	0.176	0.022	0.025	0.953
0	500	0.000	0.005	0.000	0.100	0.900
1	500	0.000	0.010	0.049	0.043	0.908
2	500	0.001	0.013	0.037	0.036	0.927
5	500	0.003	0.023	0.026	0.028	0.946
10	500	0.009	0.037	0.022	0.025	0.954
50	500	0.074	0.131	0.018	0.020	0.962

Other questions will be discussed in the next Newsletter.

PERSONAL NOTES

Our New York community will soon lose two prestigious members: Early in January of 1988, Pat Jacobs and Newton Morton will leave us to continue their career in Southampton, England. We wish them well at their new place of work.

Jurg Ott
Columbia University, Box 58
722 West 168 Street
New York, NY 10032
Tel. (212) 960-2504

Q U E S T I O N N A I R E for computer programs

If you want to share/sell your program, please fill out this
questionnaire and return it to the address at the bottom.
For packages of programs, fill out one questionnaire for each
program or, if package comprises many programs, one form only
for the whole package.

Program name (and version):

Cost (and currency if other than US$):

Runs on which machines/operating systems:

Distribution medium (floppies, which type, how many, etc.):

Source code and/or run-time version?

If source code – what programming language?

Short description:

Limitations (eg, number of observations, no consanguinity):

Data format required (particularly for family data):

Your name and address:

Please return completed questionnaire to
      Jurg Ott
      Columbia University, Box 58
      722 West 168 Street
      New York, NY 10032   USA