gene on the human Y but it is an abbreviated form and on the long arm (Craig). An interesting paper by Page suggests that the gonadoblastomas which tend to develop in dysgenetic gonads are the result of inappropriate activity of a GBY locus somewhere on Yq.

A point which clinical geneticists tend to overlook is how useful the Y is for studying overall chromosome organisation, being small and relatively unconstrained by the need for meiotic pairing or correct gene dosage. Several papers (Smith, Tyler-Smith, Cooke) describe the repetitive DNA of the telomeres, centromere, and long arm.

Most papers are detailed research reports, full of pictures of gels and lists of probes. Two more general papers at the start, however, ignore the title of the book and review sex determination in *Caenorhabditis* and *Drosophila*. I found these quite fascinating. *C elegans* normally has an XX hermaphrodite/XO male sex system based on the X:autosome ratio, but by manipulating single gene mutants and translocations it can be converted to XX female/XO male, XX female/XY male, or WZ female/ZZ male systems. Given a suitable temperature sensitive mutant, it could even be switched to alligator style sex determination by environmental temperature. There are two serious lessons from all this virtuosity. Sexual systems which appear quite different on the surface may be similar underneath. And cloning TDF is the beginning, not the end, of understanding human sex determination.

The book as a whole is a handsomely produced hardback which would grace any coffee table. It is a credit to its editors and publisher, and a standing reproach to those who produce conference papers as scruffy photoset paperbacks 18 months after the event. I recommend it to anyone who wants a detailed understanding of this strange chromosome.

ANDREW P READ

Analysis of Human Genetic Linkage
By J Ott. (Pp 223; £30·80.) Baltimore: Johns Hopkins University Press. 1985.

It is the fate of applied mathematicians to pursue a stern chase through an advancing technology. This book was published 11 years after the development and distribution of LIPED, which brought within efficient analysis the many straggling and untidy families beyond analysis by other methods.

As its title states, it is about human genetic linkage, and this is interpreted as a mathematical exercise without reference to the mechanistic concept of linkage implying membership of a linkage group, rather than mere segregational neighbourliness. No reference is made to the formal contraints imposed by direct, if sometimes erroneous, information derived from somatic cell genetics, from in situ mapping, or from the weak and uncertain networks of evidence which, with or without numerical aid, have allowed useful maps to be constructed. Not only do pedigree studies provide a declining proportion of the information on linkage, but the need for complicated methods of linkage analysis has become reduced. The DNA techniques provide codominant loci, and for such loci indirect methods of linkage analysis are often unnecessary, especially in recessive disorders.

Linkage is one of the most difficult and widely misunderstood aspects of mathematics applied to biology, and its application, often without reference to the logistic restraints of an adequate resolving power, has become a major fund user in a declining research economy; any book clarifying these difficult problems, which were hardly clarified in the paper seeding the editorial which introduced the News Genetics, is urgently needed.

Linkage at least has the advantage of there being very few papers of distinction, and even fewer books, although these few contributions are of very great distinction. This is the first substantial book since Bailey's a quarter of a century ago, a book remarkable not only for its clarity, precision, and erudition, but also for its failure to appreciate the significance both of Morton's work of six years earlier, which was the basis of almost all linkage studies until Ott's program came into general use in the late 1970s, and of the information then available from microscopy. (Much advanced mathematics was diverted to finding the mouse centromere within triads of loci, although by then it was clearly visible and outside any likely triad.) Since Morton's paper there has been little advance, apart from its translation from a sequential framework by C A B Smith and Renwick's development of computer programs for the efficient analysis of large pedigrees using likelihoods, until Ott's exploitation of an algorithm allowing inferences to be made against the direction of gamete flow in 1974, and its incorporation into his LIPED program in 1976. This has been widely used following its generous distribution and its author's help in the resolution of any difficulties, few if any of which were due to the program.

A book by the author of such an effective program could hardly fail to be a landmark. However, it is not an easy book. Indeed, it is more a collection of thinly related chapters, all of which contain information and insights of great value, although most require considerable intellectual quarrying. It is not suitable as an introduction to the

subject of either linkage or the inferences on which likelihood estimation is based.

There is no historical introduction, and the bibliography is remarkable for some omissions, including Barnard (who gave us lods), Ceppelini (who introduced the term haplotype), Fisher (only one paper, from 1935), Finney, Smith Penrose, and Smith, whose book is still one of the clearest formal expositions, and Sturtevant. The formal introduction on mechanisms is disappointing. Haplotypes are defined as a set of haploid loci derived from one gamete, although the more usual, and historically correct, usage of Ceppelini implies a set of linked loci. Map distance is introduced on a two strand model, which is needlessly at variance with reality. Centimorgans are not "percentages for recombination fractions": a centimorgan was defined by Haldane as a segment with a 1% chance of recombination. Clearly the author is clear on this, but the reader is hardly helped by a misleading plural. Interference is mentioned but not defined, or partitioned between chiasmata and chromatids, which is hardly possible without reference to a four strand model. The 22 pages of introduction lack the clarity of Bailey, whose formal introduction is undimmed by time or by recent discoveries. Ott makes short shrift of direct methods of linkage analysis by counting recombinants and non-recombinants, stating as a major disadvantage that "in practice few family pedigrees occur that would allow a direct count of recombinants and non-recombinants". This was true when LIPED was developed, but with the widespread use of codominant loci, such as restriction fragment variants, and of families selected for being informative, the need to replace the biased certainties of deduction for the uncertainties of inference is declining, or should be. The statement " . . . . searching for known recombination events in large pedigrees. Besides possible biases inherent in this technique it is generally very inefficient", clarifies the author's view. With large pedigrees and codominant loci the procedure can provide maximum likelihood estimators which are 100% efficient by definition. A system of analysis which has allowed the development of linkage maps in all other organisms mapped hardly deserves such peremptory dismissal, even if this was justified at the time of development of LIPED.

Ott makes the same point against "the tendency of some investigators to value a direct estimate more highly than an indirect estimate" even more forcibly on page 34, stating that "there is no justification for this from a statistical standpoint except that a direct estimate requires fewer observations than a comparable indirect estimate for the same precision". If direct estimates can be made on fewer observations without the need for anything so exacting as a logarithm, then there would seem the strongest justification to count the countable, and only use more complicated methods when, as is often the case, recourse must be had to inference using likelihoods.

The essential feature of Morton's approach, and Ott's computational extension of this general principle, is that exact likelihoods provide exact inferences without reference to the smallness of the numbers. However, the price of this advance is that ingrained habits based on statistical significance, normal distributions, and confidence intervals no longer apply, except in the unrealistic case of enormous sibships with loose linkage, and care needs to be taken to distinguish likelihoods by not using words with a distinct if similar connotation.

Previously, the analytical methods of maximum likelihood scores, so lucidly described by Bailey, depended on the assumption that even human sibships were sufficiently large to allow this approximation. Without computers, and the tables Morton generated from some of the earliest models, nothing more could be done. However, as Morton showed, the approximation implicit in the u scores and allied methods was far from close. Nevertheless, Ott makes considerable use of these approximations to evaluate efficiency, often to high precision. Some of the tables are difficult to follow, giving efficiencies even in the absence of linkage. Equivalent observations, which are no worse than large sample approximations (indeed they are better, since they relate to the height, rather than the curvature, of the likelihood curves) get as bad a press as the direct estimates they attempt to mimic. It is not true to state that they "have no statistically meaningful interpretation". In codominant loci in large families they contain all the information for defined male and female recombination in a simple, additive, and intelligible, if slightly biased, form. They miss out on the androgynous information of the double backcross, but this is difficult to utilise even when available, unless the sexes are merged.

The logical background to linkage is not simple, and now that major investments are being made in applying it even to non-Mendelian situations, using batteries of tests involving multiple probes, it is unfortunate that the section on likelihood does not consider these matters in detail. Morton advanced the criterion of a 1000:1 likelihood ratio (a lod of 3) as a reasonable guideline for a single marker. Clearly with 1000 markers a 'linkage' would be expected. As the numbers of tests go up the chance of error increases, but even these errors are not simple.

A statistical artefact will define a segment which

cosegregates with the locus under test, and if, as was done recently in a paper on manic depression, the same segments are tested with probes known to reside there, the error, if it is an error, will merely be rehabilitated.

There is no way in which multiplicity of probes will not, almost pro rata, increase the error rate. The whole matter is very confusing, but it is not helped by using 'significance', which refers to cumulative sets of events or areas, with 'likelihood ratios', which relate to a set of events on two hypotheses. In most cases, such as tossing a coin 10 times, while the probability of getting 10 heads and the likelihood of getting 10 heads is equivalent, for less than 10 heads it is not. The probability is 1/1024 and the likelihood ratio 1:1023. The probability of getting nine heads or more and the likelihood of getting nine heads are 11/1024 and approximately 11:1125.

Unfortunately the distinction between likelihood and probability is very difficult to clarify without likelihood distributions based on real data, and the book is devoid of these, in spite of many distributions derived from large sample approximations. A single worked example would have been helpful. The distinctive asymmetry of likelihood distributions against evenly spaced recombination fractions is of particular importance in relation to genetic prognoses, where confusion between modes and means leads to ill justified optimism on error rates.

There is a useful chapter on penetrance, well documented and illustrated by estimates based on large sample approximations. These are of major interest now that ease of entry into computer programs has led to this becoming almost habitual, even in conditions in which it can be made unnecessary by adequate clinical expertise and the exclusion of certain age ranges, as in tuberous sclerosis and neurofibromatosis. It is clearly shown that the efficiency declines drastically with close linkage if the attempt is made to 'launder' data rather than to exclude them. From these computations, although only based on large sample approximations, it would seem preferable to omit normal subjects and penetrance functions when linkage is close, as in Huntington's chorea.

This is the only substantial book on linkage since Bailey (1961), and it is likely to be the standard work on the interpretation of output from linkage programs for some time. However, it does not replace Bailey, which has a clearer formal introduction eased by typographical excellence. Nor does it replace a close reading of Morton's original paper, which provides an easier introduction to the prior expectations of linkage. Ott's program, which has been instrumental in the detection of a large number of linkages, is discussed in very limited detail. This is unfortunate as, although there are numerous print-outs on how to feed and control this robust and powerful program, it would be useful to have rather more about it, including some worked examples with likelihood distributions and their interpretation.

This is a specialised book with many valuable sections and no competitor. It was published just before the LINKAGE family of programs were in use, and these are only very briefly mentioned. At £30·80 for 200 pages it is rather expensive, but obligatory reading for any one who wishes to understand, as well as to use, likelihood methods for the estimation of recombination fractions, and who has other sources of information on the mechanism of recombination and distinctions between probability and likelihood.

J H EDWARDS