Journal of Microbiological Methods 1(1983)133–137
Elsevier

JMM 15

A simple and versatile micro-computer program for the determination of 'most probable number'

K.R. Clarke and N.J.P. Owens

Natural Environment Research Council, Institute for Marine Environmental Research, Prospect Place, The Hoe, Plymouth PL1 3DH (UK)

(Received 24 March 1983) (Accepted 10 May 1983)

Summary

A micro-computer program is described which will calculate the estimate of most probable number and an approximate confidence interval for any combination of dilution levels, numbers of replicates and sample volumes. The use of the program frees the experimenter from the constraints in experimental design often imposed by tables.

Key words: *Most probable number – Micro-computer program*

Introduction

The 'Most Probable Number' (MPN) technique is often the only method available for density estimation of specific, physiological groups of micro-organisms. Its common use involves a 10-fold dilution series with, typically, no more than five replicates at each dilution; this leads to very wide confidence intervals for the estimate of population density and a widespread belief that the method is too imprecise to be of value. A more reliable MPN can be obtained by increasing the number of replicates and decreasing the size of dilution, but the maintenance of a constant dilution factor and fixed numbers of replications, throughout the series, will usually involve prohibitive levels of experimentation. Yet this restriction to a symmetrically designed dilution series is an artificial constraint, imposed by the need to refer to existing MPN tables. Many of the additional replicates and dilution levels provide no worthwhile information, improvement in accuracy following mainly from increased sampling at the 'critical' dilution levels (where neither the positive nor the negative scores are zero).

No set of MPN tables can be truly comprehensive, even for a balanced series, and it is clearly impossible to construct general tables which cater for the more relevant design strategy of increased sampling at the critical dilution levels. However, a computer algorithm is not constrained in this way; we therefore present a program, designed for easy implementation on a micro-computer, which will calculate the estimate of MPN (and an approximate confidence interval) for *any* combination of dilution levels, numbers of replicates and sample volumes. We suggest that the increasing availability of personal computers makes tabulation redundant but, if desired, this program can be used to construct specific MPN tables, for a balanced series involving a non-standard dilution factor, for example.

Theory

Denote the mean number of bacteria per unit volume, at the initial dilution level, by λ (the MPN). There are m dilution *levels* $k_1, \ldots, k_m$ (in standard trials $k_i = k^{i-1}$, where k is the dilution *factor*), the ith level involving n_i replicates, each with sample volume V_i. Under the usual assumption that the bacteria are randomly dispersed, the number of positive scores at the ith level, x_i, has a Binomial (n_i, p_i) distribution, with $p_i = 1 - \exp(-\lambda V_i/k_i)$.

Estimation of λ is by maximization of the log likelihood (as a function of log λ) based on the scores from *all m* levels; this is achieved by a Newton-Raphson iteration [1], employing analytic derivatives. Except for some degenerate cases, a starting value for λ is obtained as the slope from a regression (through the origin) of $-\log(1-x_i/n_i)$ on (V_i/k_i). This procedure is usually stable and convergence is quick, though not guaranteed if the data are clearly inconsistent with the model assumptions. A confidence interval for λ is calculated by approximating the log likelihood function, in the region of its maximum, by a quadratic in log λ.

The Program

A computer program to perform the MPN estimation is given in Fig. 1. It is written in fairly elementary BASIC and will run with little or no modification on most micro-computers. Fig. 2 gives two example runs, contrasting the options of a standard trial with a more flexible design, where larger numbers of replicates and a lower dilution factor are used at the critical dilution level, to improve the accuracy of estimation. Note that the scores from *all* dilutions should be entered; the data usually ignored in construction of the 'numerical characteristic', for entry to standard tables (e.g. ref. 2), *can* be relevant to the estimation, particularly for small dilution factors. The resulting estimate is always the MPN at the first dilution level entered (usually 1).

It is advisable to check that the 'model scores', predicted for each level from the estimated MPN, are in rough agreement with those observed. A formal test of the model assumptions could be constructed, from a generalised likelihood ratio

```
  10 PRINT'MPN ESTIMATION FOR GENERAL DILUTION TRIAL':PRINT
 100 REM   UP TO 50 DILUTION LEVELS
 101 DIM C(50),K(50),N(50),P(50),X(50),Y(50),Z(50)
 150 REM   A0 IS LARGEST VALUE ALLOWED IN EXP(A0) AND EXP(-A0)
 151 REM   ESTIMATES CALCULATED TO 4 SIGNIFICANT FIGURES
 152 A0=86:T0=1.0E-4
 199 REM -------------------------------------------------------------
 200 REM   INPUT DATA
 210 INPUT 'HOW MANY DILUTION LEVELS';M
 300 INPUT'SAME NUMBER OF REPLICATES AT EACH LEVEL (Y/N)';Q1$
 310 IF Q1$='N' THEN GOTO 330
 320 INPUT'HOW MANY';N0:FOR I=1 TO M:N(I)=N0:NEXT I:GOTO 350
 330 PRINT'ENTER NUMBER OF REPLICATES ONE BY ONE'
 340 FOR I=1 TO M:INPUT N(I):NEXT I
 350 INPUT'SAME DILUTION FACTOR AT EACH LEVEL (Y/N)';Q2$
 360 IF Q2$='N' THEN GOTO 380
 370 INPUT 'WHAT IS IT';K:FOR I=1 TO M:K(I)=K^(I-1):NEXT I:GOTO 400
 380 PRINT'ENTER DILUTION LEVELS (NOT FACTORS!) ONE BY ONE'
 390 FOR I=1 TO M: INPUT K(I):NEXT I
 400 INPUT 'SAME VOLUME USED AT EACH LEVEL (Y/N)';Q3$
 410 IF Q3$='N' THEN GOTO 430
 420 INPUT 'WHAT IS IT';V:FOR I=1 TO M:C(I)=V/K(I):NEXT I:GOTO 450
 430 PRINT'ENTER VOLUMES USED ONE BY ONE'
 440 FOR I=1 TO M:INPUT V:C(I)=V/K(I):NEXT I
 450 PRINT'ENTER NUMBER OF POSITIVE RESPONSES AT EACH LEVEL, ONE BY ONE'
 460 FOR I=1 TO M:INPUT X(I):NEXT I
 499 REM -------------------------------------------------------------
 500 REM   CALCULATE STARTING VALUES FOR ITERATION
 501 G=0:G1=0:G2=0:G3=0
 510 FOR I=1 TO M:IF X(I)=N(I) THEN G2=G2+1:G1=I:GOTO 540
 520 IF X(I)=0 THEN G3=I:GOTO 540
 530 G=G+1:Y(G)=-LOG(1-X(I)/N(I)):Z(G)=C(I)
 540 NEXT I
 550 IF G>0 THEN GOTO 600
 560 IF G2=0 THEN PRINT'NO POSITIVES. MPN=0':GOTO 1000
 570 IF G2=M THEN PRINT'ALL POSITIVES. MPN=INFINITY':GOTO 1000
 580 IF G1=M THEN C1=C(G3+1):C2=C(G3):N1=N(G3+1):N2=N(G3):GOTO 590
 585 C1=C(G1):C2=C(G1+1):N1=N(G1):N2=N(G1+1)
 590 M0=LOG(LOG((C1*N1+C2*N2)/(C2*N2))/C1):GOTO 700
 600 S1=0:S2=0
 610 FOR I=1 TO G:S1=S1+Y(G)*Z(G):S2=S2+Z(G)*Z(G):NEXT I
 620 M0=LOG(S1/S2)
 699 REM -------------------------------------------------------------
 700 REM   NEWTON-RAPHSON ITERATION
 701 J=0
 710 J=J+1:E=EXP(M0):D1=0:D2=0
 720 FOR I=1 TO M:E0=C(I)*E:P(I)=1:IF E0<A0 THEN P(I)=1-EXP(-E0)
 730 IF P(I)=0 THEN GOTO 760
 740 D1=D1+C(I)*X(I)/P(I)-N(I)*C(I)
 750 D2=D2+C(I)*C(I)*N(I)*(1-P(I))/P(I)
 760 NEXT I
 770 D1=D1*E:D2=D2*E^2:M1=M0+D1/D2
 780 IF J>50 OR ABS(M1)>A0 THEN PRINT'DIVERGENCE. CHECK DATA':GOTO 1000
 790 IF ABS(M1-M0)>T0*ABS(M0) THEN M0=M1:GOTO 710
 799 REM -------------------------------------------------------------
 800 REM   OUTPUT RESULTS
 801 PRINT:PRINT'CONVERGENCE IN';J;'ITERATIONS'
 810 S=SQR(1/D2):L1=M1-2*S:L2=M1+2*S:IF L2>A0 THEN L2=A0
 820 L=EXP(M1):L1=EXP(L1):L2=EXP(L2)
 830 PRINT'MPN / UNIT VOLUME =';L
 840 PRINT'APPROXIMATE 95% CONFIDENCE INTERVAL = (';L1;',';L2;')'
 850 PRINT:PRINT'DILUTION','REPLICATES','SCORE','MODEL SCORE'
 860 FOR I=1 TO M:PRINT K(I),N(I),X(I),N(I)*P(I):NEXT I
1000 END
```

Fig. 1. Listing of the 'Most Probable Number' program, written in the BASIC language.

a

```
RUN

MPN ESTIMATION FOR GENERAL DILUTION TRIAL

HOW MANY DILUTION LEVELS? 5
SAME NUMBER OF REPLICATES AT EACH LEVEL (Y/N)? Y
HOW MANY? 5
SAME DILUTION FACTOR AT EACH LEVEL (Y/N)? Y
WHAT IS IT? 10
SAME VOLUME USED AT EACH LEVEL (Y/N)? Y
WHAT IS IT? 1
ENTER NUMBER OF POSITIVE RESPONSES AT EACH LEVEL, ONE BY ONE
? 5
? 5
? 3
? 0
? 0

CONVERGENCE IN 3 ITERATIONS
MPN / UNIT VOLUME = 78.2001
APPROXIMATE 95% CONFIDENCE INTERVAL = ( 25.188 , 242.784 )

DILUTION          REPLICATES        SCORE             MODEL SCORE
 1                5                 5                 5
 10               5                 5                 4.998
 100              5                 3                 2.71321
 1000             5                 0                 .376236
 10000            5                 0                 .0389618
```

b

```
READY

RUN

MPN ESTIMATION FOR GENERAL DILUTION TRIAL

HOW MANY DILUTION LEVELS? 5
SAME NUMBER OF REPLICATES AT EACH LEVEL (Y/N)? N
ENTER NUMBER OF REPLICATES ONE BY ONE
? 3
? 3
? 10
? 10
? 5
SAME DILUTION FACTOR AT EACH LEVEL (Y/N)? N
ENTER DILUTION LEVELS (NOT FACTORS!) ONE BY ONE
? 1
? 10
? 100
? 200
? 400
SAME VOLUME USED AT EACH LEVEL (Y/N)? N
ENTER VOLUMES USED ONE BY ONE
? 1.1
? 1.05
? .95
? .88
? 1
ENTER NUMBER OF POSITIVE RESPONSES AT EACH LEVEL, ONE BY ONE
? 3
? 3
? 8
? 3
? 0

CONVERGENCE IN 3 ITERATIONS
MPN / UNIT VOLUME = 105.486
APPROXIMATE 95% CONFIDENCE INTERVAL = ( 57.6162 , 193.165 )

DILUTION          REPLICATES        SCORE             MODEL SCORE
 1                3                 3                 3
 10               3                 3                 2.99985
 100              10                8                 8.33044
 200              10                3                 4.03745
 400              5                 0                 1.15944


READY
```

Fig. 2. Examples of the use of the program: (a) for a 10- dilution series at five levels, with constant replication numbers and sample volumes, (b) for an asymmetric dilution series at five levels, with variable replication numbers and sample volumes.

or χ^2 statistic [1], but there are rarely sufficient replicates to validate the χ^2 approximation.

Conclusions

Use of this program frees the experimenter from the artificial constraints (equal replication, equal dilution factors) imposed by MPN tables, allowing the allocation of additional replicates, and interpolated dilutions, as the critical levels are encountered. In addition, prior assessment of the efficacy of carrying out extra trials, in terms of reduction of confidence interval widths, can be made by running the program for the planned design and for likely outcomes (predicted from current data).

Acknowledgements

This work forms part of the research programme of the Institute for Marine Environmental Research, a component body of the Natural Environment Research Council.

References

1 Silvey, S.D. (1975) Statistical Inference. Chapman and Hall, London.
2 Rodina, A.G. (1972) Methods in Aquatic Microbiology. Translated by R.R. Colwell and M.S. Zambruski. Butterworths, London, p. 179.