N86-24544

SIMULATION AND STUDY OF SMALL NUMBERS OF RANDOM EVENTS

BY

RUSSELL D. SHELTON

INSTRUCTOR IN PHYSICS ELECTRONICS AND DATA PROCESSING
J. C. CALHOUN COMMUNITY COLLEGE
DECATUR, ALABAMA

ABSTRACT

RANDOM EVENTS WERE SIMULATED BY COMPUTER AND SUBJECTED TO VARIOUS STATISTICAL METHODS TO EXTRACT IMPORTANT PARAMETERS. VARIOUS FORMS OF CURVE FITTING WERE EXPLORED, SUCH AS LEAST SQUARES, LEAST DISTANCE FROM A LINE, MAXIMUM LIKELIHOOD. PROBLEMS CONSIDERED WERE DEAD TIME, EXPONENTIAL DECAY, AND SPECTRUM EXTRACTION FROM COSMIC RAY DATA USING BINNED DATA AND DATA FROM INDIVIDUAL EVENTS. COMPUTER PROGRAMS, MOSTLY OF AN ITERATIVE NATURE, WERE DEVELOPED TO DO THESE SIMULATIONS AND EXTRACTIONS AND ARE PARTIALLY LISTED AS APPENDICES. THE MATHEMATICAL BASIS FOR THE COMPUTER PROGRAMS IS GIVEN IN THE TEXT OF THE REPORT.

ACKNOWLEDGEMENTS

DR. P. B. EBY SUGGESTED THE POSSIBLE INTEREST IN EXTRACTING DECAY PARAMETERS FROM RARE EVENTS. A NOTE BY DR. R. F. ELSNER STIMULATED A LOOK AT DEAD TIME CORRECTIONS. CONVERSATIONS WITH DRS. R. DECHER, E. URBAN, P. PETERS, AND T. PARNELL HELPED WITH THE CHOICE OF WORK TO BE DONE. MR. JOHN WATTS PROVIDED COSMIC RAY DATA AND HELPFUL INFORMATION ON HOW IT HAD BEEN PROCESSED AND INTERPRETED.

TABLE OF CONTENTS

SIMULATION AND STUDY OF SMALL NUMBERS OF RANDOM EVENTS

2. RELATIONSHIP OF BINOMIAL, POISSON, AND NORMAL DISTRIBUTIONS

IF WE DIVIDE A TIME INTERVAL T INTO n EQUAL SMALL SUBINTERVALS OF LENGTH Dt, A RANDOM EVENT IS EQUALLY LIKELY TO OCCUR DURING ANY SUBINTERVAL OF DURATION Dt. THE PROBABILITY Dp OF AN EVENT DURING ANY GIVEN SUBINTERVAL Dt IS GIVEN BY THE EXPRESSION,

Dp = L Dt , 2.1

WHERE L IS THE AVERAGE NUMBER OF EVENTS PER UNIT TIME AND Dt IS SO SMALL THAT TWO COUNTS DURING THE INTERVAL Dt IS HIGHLY UNLIKELY. THE PROBABILITY THAT AN EVENT WILL NOT OCCUR IN ANY GIVEN INTERVAL Dt IS GIVEN BY THE EXPRESSION,

1 - Dp = 1 - L Dt 2.2

THE PROBABILITY P(N,T) OF N EVENTS IN TIME T IS GIVEN BY THE BINOMIAL DISTRIBUTION,

P(N,T) = [L*Dt]^N * [(1-Dt L)^(n-N)]* n!/[(n-N)! N!] . 2.3

IN EQ 2.3 AND FOLLOWING, THE SYMBOL * DENOTES MULTIPLICATION AND THE SYMBOL ^ PRECEDES AN EXPONENT. P(N,T) IS THE Nth TERM OF THE POLYNOMIAL [L*Dt+(1-L*Dt)]^n. THIS DISTRIBUTION IS THEREFORE NORMALIZED, AND ANY DISTRIBUTION DERIVED FROM IT IN A LIMITING PROCESS SHOULD BE NORMALIZED.

THE POISSON DISTRIBUTION IS EASILY OBTAINED FROM THE BINOMIAL DISTRIBUTION BY USING THE STERLING APPROXIMATION,

LOG n! ~ (1/2) LOG (2 PI) + (n + 1/2) LOG n - n , 2.4

AND APPROXIMATIONS LIKE

LOG (n-N) = LOG [n (1 - N / n)] ~ LOG n - n/N . 2.5

THESE APPROXIMATIONS ARE QUITE GOOD WHEN n >> N. IF WE TAKE THE NATURAL LOGARITHM OF EQ 2.3 AND USE EQUATIONS 2.4 AND 2.5, WE OBTAIN THE POISSON DISTRIBUTION,

P(N,T) = [(L T) N]*[EXP (-L*T)]/N! 2.6

THE AVERAGE NUMBER OF EVENTS N IN TIME T IS (L*T) AND THE STANDARD DEVIATION FOR THE POISSON DISTRIBUTION IS SQR (L*T).

AS THE NUMBER OF EVENTS N BECOMES LARGE, THE POISSON BECOMES EQUIVALENT TO THE NORMAL DISTRIBUTION. IF WE TAKE THE NATURAL LOGARITHM OF THE POISSON DISTRIBUTION AND USE THE STIRLING APPROXIMATION AND THE APPROXIMATION,

LOG [1+X/(L*T)] = X / L T - [(X/L*T)^2] / 2 , 2.7

WHERE

$$X = N - L T , \qquad 2.8$$

WE FIND THAT

$$P(X,T) \sim SQR [1 / (2 PI L*T] * EXP [-X^2/(2 L*T)], \qquad 1.9$$

WHICH IS THE NORMAL DISTRIBUTION WITH AVERAGE LT (X = 0) AND STANDARD DEVIATION SQR (L T). TABLE 2.1 COMPARES THE BINOMIAL, POISSON AND NORMAL DISTRIBUTIONS FOR N = 30 , T = 5 , AND L = 3 .

3 DEAD TIME CORRECTIONS FOR RANDOM EVENTS.

SUPPOSE THAT AN EVENT DETECTOR (COUNTER) HAS A DEAD TIME t AFTER EACH COUNT. IF N' RANDOM EVENTS ARE OBSERVED DURING TIME T, WE KNOW THAT THE COUNTER WAS DEAD (INOPERATIVE) FOR A TIME N'*t. THE AVERAGE NUMBER OF EVENTS EXPECTED DURING THIS DEAD TIME WOULD BE L*N'*t AND WOULD ALSO OBEY A POISSON DISTRIBUTION. THE NUMBER OF EVENTS N IS ESTIMATED TO BE GIVEN BY THE EQUATION,

$$N = N' + L N' t \qquad . \qquad 3.1$$

IT DOES NOT MATTER HOW OFTEN OUR COUNTER IS TURNED ON OR OFF. IF THE EVENTS ARE RANDOM AND L IS A CONSTANT, EQ 3.1 IS VALID WHEN N' REPRESENTS ALL THE OBSERVED COUNTS AND T REPRESENTS THE TOTAL TIME THAT THE COUNTER IS ON. IF WE MAKE THE ASSUMPTION THAT N IS EQUAL TO L*T , EQ 3.1 BECOMES

$$L*T = N' + L*N'*t \qquad 3.2$$

OR

$$L = N'/ (T - N'*t) \qquad . \qquad 3.3$$

EQ 3.3 IS THE SIMPLE STATEMENT THAT THE BEST ESTIMATE OF THE NATURAL COUNT RATE L IS THE TOTAL NUMBER OF COUNTS DIVIDED BY THE TOTAL TIME THE COUNTER IS ON. IT IS POOR PRACTICE TO REPLACE L IN EQ 3.1 BY N / T AND USE THE RESULT,

$$N = N' / (1 - N' t / T), \qquad 3.4$$

AS A CORRECTED COUNT. THIS PRACTICE OVERCORRECTS WHEN N' IS LARGE AND UNDERCORRECTS WHEN N' IS SMALL.

THERE ARE A NUMBER OF TECHNIQUES WHICH MAY BE USED TO STUDY COUNTER DEAD TIME. A FIRST APPROACH IS TO OVERWHELM THE COUNTER WITH MANY MORE EVENTS THAN IT CAN POSSIBLY COUNT. IF THE EVENT RATE L IS RELATED TO THE COUNT RATE R BY THE EQUATION,

$$R = L / (1 + L*t) \qquad , \qquad 3.5$$

WHERE t IS THE DEAD TIME AFTER EACH EVENT, THE MAXIMUM POSSIBLE COUNT RATE R(MAX) IS SEEN TO BE GIVEN AS 1 / t, SO THAT THE DEAD TIME t = 1 / R(MAX). IF THE COUNTER DETECTS RADIATION, A MORE

TABLE 2.1

A comparison of binomial, Poisson, and Gaussian distributions for n = 30, T = 5, and $\lambda = 3$ corresponding to an average count of 15

N	B(N)	P(N)	G(N)
1.000	0.000	0.000	0.000
2.000	0.000	0.000	0.000
3.000	0.000	0.000	0.001
4.000	0.000	0.001	0.002
5.000	0.000	0.002	0.004
6.000	0.001	0.005	0.007
7.000	0.002	0.010	0.012
8.000	0.005	0.019	0.020
9.000	0.013	0.032	0.031
10.000	0.028	0.049	0.045
11.000	0.051	0.066	0.060
12.000	0.081	0.083	0.076
13.000	0.112	0.096	0.090
14.000	0.135	0.102	0.100
15.000	0.144	0.102	0.103
16.000	0.135	0.096	0.100
17.000	0.112	0.085	0.090
18.000	0.081	0.071	0.076
19.000	0.051	0.056	0.060
20.000	0.028	0.042	0.045
21.000	0.013	0.030	0.031
22.000	0.005	0.020	0.020
23.000	0.002	0.013	0.012
24.000	0.001	0.008	0.007
25.000	0.000	0.005	0.004
26.000	0.000	0.003	0.002
27.000	0.000	0.002	0.001
28.000	0.000	0.001	0.000
29.000	0.000	0.000	0.000
30.000	0.000	0.000	0.000

SUITABLE APPROACH INITIALLY IS TO USE A RADIATION SOURCE AND THE INVERSE SQUARE LAW FOR CALIBRATION. THE EVENT RATE AT THE COUNTER IS GIVEN AS

$$R = K / r^2 \quad , \qquad 3.6$$

WHERE K IS PROPORTIONAL TO THE SOURCE INTENSITY. FROM EQS 3.5 AND 3.6,

$$R = L / (1 + L*T) = 1 / [(r^2 / K) + t], \qquad 3.7$$

SO THAT

$$R*r^2 = - R*K*t + K \quad . \qquad 3.9$$

A PLOT OF R*r 2 VERSUS R SHOULD BE A STRAIGHT LINE WITH SLOPE K * AND INTERCEPTS,

$$R*r^2 = K \quad ; \quad R = 1 / t \; . \qquad 3.10$$

IF THIS IS NOT SO, DEAD TIME t IS NOT INDEPENDENT OF COUNT RATE.

4. COMMON STATISTICS FOR EXPONENTIAL DECAY

THE PROBABILITY Dp OF A RADIOISOTOPE DECAY DURING THE TIME INTERVAL DT IS GIVEN BY THE EQUATION,

$$Dp = L*N*DT \quad , \qquad 4.1$$

WHERE N IS THE NUMBER OF ATOMS AVAILABLE FOR DECAY AND L IS THE DECAY CONSTANT. IF WE DIVIDE TIME T INTO n EQUAL INTERVALS DT, THE PROBABILITY OF NO DECAY IN TIME T FOLLOWED BY A DECAY IN TIME DT IS GIVEN AS

$$P(T)*Dt = (1 - L*N*DT)^n *L*N*DT \quad . \qquad 4.2$$

AS n BECOMES LARGE THIS REDUCES TO

$$P(T)*dT = [EXP (- L*N*T)]*L*N*dT \quad . \qquad 4.3$$

THIS DISTRIBUTION OF THE TIME BETWEEN DECAYS HAS AN AVERAGE OF L*N AND A STANDARD DEVIATION OF L*N. GIVEN THE TIMES T(I) ASSOCIATED WITH A NUMBER OF DECAYS N(I), A NUMBER OF TECHNIQUES HAVE BEEN USED TO COMPUTE THE ORIGINAL NUMBER OF ATOMS N(O) AND THE DECAY CONSTANT L. IF THERE ARE MANY DECAYS PER UNIT TIME, IT IS CONVENIENT TO WRITE THE DIFFERENTIAL EQUATION,

$$dN / dT = - L N \quad , \qquad 4.4$$

WHICH HAS THE SOLUTION,

$$N(T) = N(O) EXP (- L T) \; . \qquad 4.5$$

SINCE THE COUNT RATE IS PROPORTIONAL TO N(T), A PLOT OF LOG COUNT RATE VERSUS TIME HAS A SLOPE OF -L AND AN INTERCEPT L*N(O).

IF EVENTS ARE FEW, WE NEED TO DEVELOP TECHNIOUES TO EXTRACT AS MUCH INFORMATION AS POSSIBLE FROM THE DATA AT HAND. TO TEST THE VARIOUS TECHNIOUES, DECAY DATA WAS GENERATED BY COMPUTER AND THE DATA WAS PROCESSED BY EACH TECHNIOUE TO SEE HOW WELL IT WOULD DO.

A COMPUTER GENERATES RANDOM NUMBERS X(I) DISTRIBUTED UNIFORMLY IN THE INTERVAL BETWEEN ZERO AND UNITY. TO PRODUCE NUMBERS CORRESPONDING TO RANDOM DECAY TIMES, WE EOUATE THE DISTRIBUTIONS,

$$P(T)\ dT = F(X)\ dX, \qquad 4.6$$

WHERE P(T) COMES FROM EO 4.3 AND F(X) IS UNITY. THE INDEFINITE INTEGRAL OF BOTH SIDES OF EO 4.6 YIELDS

$$T = -\ (\ 1\ /\ N*L\ )\ LOG\ (\ 1 - X\ )\ . \qquad 4.7$$

BECAUSE X IS A RANDOM VARIABLE, WE CAN WRITE

$$T(I) = -\ (\ 1\ /\ N(I)*L\ )\ LOG\ (\ 1 - X(I)\ )\ , \qquad 4.8$$

WHERE

$$N(I) = N(O) - I + 1\ . \qquad 4.9$$

IF EVENTS ARE FEW, WE CAN RECORD THE TIME T(I) FOR EACH DECAY AND FIT EO 4.5 TO THE DATA. THE RESULTS ARE GENERALLY POOR IF WE USE LEAST SOUARES FITS. USING THE DISTRIBUTION OF EO 4.3, WE CAN CALCULATE THE AVERAGE TIME t(I) BETWEEN DECAYS TO BE

$$t(I) = 1\ /\ [\ (\ N(O) - I + 1)*L\ ]\ . \qquad 4.10$$

AFTER REARRANGING,

$$I*t(I) = (\ N(O) + 1\ )*t(I) - 1\ /\ L\ . \qquad 4.11$$

A LEAST SOUARES CURVE FIT OF I*t(I) VERSUS t(I) SHOULD YIELD A SLOPE OF N(O)+1 AND AN INTERCEPT OF -1 / L . THIS OFFERS LITTLE IMPROVEMENT. WE COULD REARRANGE EO 4.5 TO THE FORM,

$$T(I) = (1/L)\ LOG\ N(O) - (1/L)\ LOG\ N(I)\ . \qquad 4.12$$

A LEAST SOUARES FIT OF T(I) VERSUS LOG N(I) YIELDS A SLOPE OF -1/L AND AN INTERCEPT OF (1/L) LOG N(O).

WE MIGHT THINK THAT THE METHOD OF LEAST SOUARES GIVES TOO MUCH WEIGHT TO ERRATIC (RARE) EVENTS AND THAT SOME SCHEME TO GIVE LESS WEIGHT TO POINTS DISTANT FROM THE CURVE FIT MIGHT BE BETTER. A SOUARE ROOT AND FOURTH ROOT OF THE SUM OF SOUARES RESULTED IN SMALL IMPROVEMENTS.

AS A FINAL EXERCISE, THE METHOD OF MAXIMUM LIKELIHOOD WAS USED TO DERIVE L AND N(0). IT IS ASSUMED THAT A SEQUENCE OF DECAY EVENTS, DESCRIBED BY I AND t(I), THE TIME BETWEEN EVENTS I AND I-1, HAS A PROBABILITY CLOSE TO THE MAXIMUM. THE PROBABILITY FOR SUCH A SEQUENCE CAN BE WRITTEN AS

P[t(1),t(2),...] = F*P(1)*P(2)*P(3)*......   4.13

WHERE F IS A CONSTANT OF PROPORTIONALITY AND

P(I) = L*N(I)*EXP[-N(I)*L*t(I)]   .   4.14

THIS PRODUCT IS MORE MANAGEABLE AS A LOGARITHM IN THE FORM,

LOG F*P[t(1),t(2),...]=SUM(I){LOG[L*N(I)]-N(I)*L*t(I)} 4.15

USING AN ITERATION SCHEME WE CHOOSE VALUES OF N(0) AND L WHICH MAXIMIZE THIS FUNCTION.

RESULTS USING THE VARIOUS TECHNIQUES ARE SHOWN IN TABLE 4.1. IT IS SEEN THAT THE METHOD OF MAXIMUM LIKELIHOOD IS CONSISTENTLY SUPERIOR FOR N(0). ... PLACING LESS EMPHASIS ON DISTANT POINTS IS WORTHWHILE IF THE METHOD OF MAXIMUM LIKELIHOOD IS NOT USED. TABLE 4.1 SHOWS THE RESULTS OF USING THE DIFFERENT METHODS ON THE SAME SET OF DATA.

AS AN ASIDE, IT IS NOTED THAT EQ 4.15 COULD BE USED FOR A POISSON DISTRIBUTION IF L*N(T) IS REPLACED BY L. THE RESULT IS GIVEN AS

LOG P[t(1),t(2),...] = SUM(I) LOG[L*EXP(-L*t(I))]   ,
= N*LOG L - L*T   ,   4.16

WHERE T IS THE TOTAL TIME FOR N COUNTS. THE MAXIMIZATION OF THIS FUNCTION OCCURS WHEN L=N/T, SO THAT THE DISTRIBUTION OF TIME BETWEEN EVENTS IS NOT USEFUL IN STUDYING A POISSON DISTRIBUTION. ONLY THE TOTAL TIME T AND THE TOTAL COUNTS N IS IMPORTANT.

5. TREATMENT OF POWER SPECTRUM DATA

EVENTS OBEYING THE DISTRIBUTION,

P(E)*dE = A*E^-B*dE ,   5.1

ARE CHARACTERISTIC OF COSMIC RAYS AND RADIATION BELTS, WHERE E IS THE PARTICLE ENERGY AND P(E)*dE IS THE PROBABILITY OF OBSERVING A PARTICLE WITH ENERGY BETWEEN E AND E+dE. THE NORMALIZATION OF THIS FUNCTION RESULTS IN THE EQUALITY,

A = (B-1)/(LL^(-B+1)-UL^(-B+1))   ,   5.2

WHERE LL IS THE LOWER ENERGY LIMIT AND UL IS THE UPPER ENERGY LIMIT. USING THE METHOD OF SECTION 4, THIS SPECTRUM IS SIMULATED

TABLE 4.1

SUMMARY OF FINDINGS ON CURVE FITTING

N(O) = 5 LAMBDA = 1 SEQUENCES = 7

METHOD#	LAMBDA	S. D.	N(O)	S. D.
1	2.45	5.23	3.68	1.14
2	0.71	0.58	5.46	1.62
3	1.01	0.62	6.13	2.66
4	1.24	1.08	7.49	6.61

N(O) = 10 LAMBDA = 1 SEQUENCES = 7

METHOD#	LAMBDA	S. D.	N(O)	S. D.
1	2.03	1.08	8.56	0.69
2	0.93	0.50	10.65	2.27
3	0.99	0.52	11.32	2.50
4	0.77	0.18	10.46	1.19
5	1.04	0.28	10.06	1.96

N(O) = 30 LAMBDA = 1 SEQUENCES = 7

METHOD#	LAMBDA	S. D.	N(O)	S. D.
1	2.03	0.45	27.92	1.44
2	0.93	0.07	28.21	4.05
3	0.99	0.08	29.65	4.30
4	1.00	0.12	30.11	0.29

N(O) = 50 LAMBDA = 1 SEQUENCES = 7

METHOD#	LAMBDA	S. D.	N(O)	S. D.
1	1.39	0.22	47.02	1.67
2	0.91	0.14	47.12	3.01
3	0.93	0.14	48.06	2.87
4	1.00	0.11	50.51	0.81
5	1.01	0.11	50.46	1.00
6				

METHODS

1. L.S.F. TO I*TAU(J) = (N(O) + 1)*TAU(1) - 1/L
2. L.S.F. TO LOG (N(O) - I) = -L*T(I) + LOG (N(O))
3. L.S.F. TO T(I) = (-1/L)*LOG (N(O) - I) + (1/L) LOG (N(O))
4. MAXIMUM LIKELIHOOD ITERATION
5. LEAST DISTANCE ITERATION
6. ROOT MEAN SQUARE ITERATION

L.S.F. - LEAST SQUARES FIT ; L - DECAY RATE CONSTANT
TAU(I) - TIME BETWEEN EVENTS I-1 AND I ; N(O) - ORIGINAL
NUMBER OF EVENTS ; T(I) - TIME WHEN EVENT I OCCURS.

TABLE 4.2

COMPARISON OF SOME METHODS USING THE SAME SET OF DATA

SEED = 11111 EVENTS = 30 LAMBDA = 1

N(0)	LAMBDA	METHOD*	EQUATION**
28.9	1.27	1	1
27.5	0.88	1	2
28.8	0.93	1	3
29.3	2.05	2	1
29.7	0.97	2	2
29.8	0.78	2	3
29.1	2.99	3	1
28.9	0.76	3	2
33.9	1.17	3	3
29.2	1.99	4	1
27.5	0.88	4	2
30.6	1.01	4	3
29.5	1.14	5	4

* METHODS
1. LEAST SUM OF SQUARES OF (Y(I)-M*X(I)-B)
2. LEAST SUM OF ABS(Y(I)-M*X(I)-B)
3. LEAST SUM OF SQRT(ABS(Y(I)-M*X(I)+B))
4. LEAST SUM OF DISTANCES OF POINTS FROM LINE
5. MAXIMUM LIKELIHOOD

** EQUATIONS
1. I*t(I) = (N(0)+1)*t(I)-1/L
2. LOG N(I) = -L*T(I) + LOG N(0)
3. T(I) = -[LOG N(I)]/L + [LOG N(0)]/L

N(I) = EVENTS REMAINING
T(I) = TIME WHEN Ith EVENT OCCURS
t(I) = T(I)-T(I-1)

BY THE RANDOM NUMBER STATEMENT,

E(I) = (LL^(-B+1) - [(B-1)*X(I)/A])^(1/(B-1)) , 5.3

WHERE X(I) IS A RANDOM NUMBER DISTRIBUTED UNIFORMLY IN THE INTERVAL [0,1]. THE MAXIMUM LIKELIHOOD STATEMENT IS WRITTEN AS

LOG P(1,2,..) = SUM(I) [LOG A - B*LOG E(I) . 5.4

TABLE 5.1 SHOWS THE EFFECTIVENESS OF THE MAXIMUM LIKELIHOOD METHOD OF EXTRACTING THE EXPONENT B FOR THE DISTRIBUTION.

IF DATA IS BINNED, THE PROBABILTY P [E(I),E(I+1)] OF AN AN OBSERVED PARTICLE HAVING ENERGY IN THE "BIN" BETWEEN E(I) AND E(I+1) IS THE INTEGRAL OF THE NORMALIZED DISTRIBUTION FUNCTION BETWEEN THESE ENERGIES OR

P[E(I),E(I+1)] = (A/(B 1))[E(I) (-B+1) E(I+1) (-B+1)]. 5.5

THE MAXIMUM LIKELIHOOD STATEMENT FOR A BINNED DATA IS

LOG P [1,2,3,...] - SUM(I) P(I) LOG (O) 5.6

WHERE

O=([E(I) (B+1)-E(I+1) (-B+1)]/[LL (-B+1)-UL (-B+1)]. 5.7

THE RESULTS OF SOME COMPUTER EXTRACTION OF EXPONENTS FOR POWER LAW SPECTRA ARE GIVEN IN TABLE 5.1. IT SHOULD BE NOTED THAT BINNING IS A WAY OF INTRODUCING UNCERTAINTIES IN THE DATA OR A WAY OF THROWING AWAY INFORMATION EITHER BECAUSE OF CONVENIENCE OR NECESSITY. THE RESULTS OF TABLE 5.2 ARE FOR BINS OF EQUAL SIZE. THE ENERGIES E(I) DEFINING BINS COULD HAVE BEEN CHOSEN BY GEOMETRIC PROGRESSION OR INTERVALS OF EQUAL PROBABILITY OR ANY OTHER WAY DESIRED.

AS A SUMMARY, WE CAN SAY THAT DATA SHOULD NOT BE BINNED UNLESS IT IS INESCAPABLE.

6. RANDOM ERROR AND SPECTRUM EXTRACTION

FOR VARIOUS REASONS, THE RESPONSE OF THE DETECTING INSTRUMENTS IS DIFFERENT FROM THE REAL ENERGY SPECTRUM. THE LEAST TROUBLESOME ERRORS ARE THOSE WHICH RESULT IN AN ENERGY UNCERTAINTY OVER A LIMITED RANGE. BINNING, FOR EXAMPLE, INTRODUCES AN UNCERTAINTY EQUAL TO THE BIN WIDTH, AND WE CAN SIMULATE QUITE EASILY WHAT BINNING DOES TO THE SPECTRUM EXTRACTION. GIVEN A BIN DISTRIBUTION AND THE ABILITY TO SIMULATE THE EFFECT OF BINNING ON A GIVEN SPECTRUM, WE CAN ALWAYS FIND A SPECTRUM WHICH REPRODUCES, WITHIN THE VAGARIES OF STATISTICS, THE BIN DISTRIBUTION. SOME INSTRUMENTS WILL DETECT PARTICLES OF A GIVEN ENERGY E(I) AND INDICATE A DISTRIBUTION OF ENERGIES E(J) SUCH THAT 0 < E(J) < E(I). THE PROBLEM OF EXTRACTING THE ORIGINAL SPECTRUM THEN BECOMES DIFFICULT OR IMPOSSIBLE.

TABLE 5.1

MAXIMUM LIKELIHOOD TREATMENT OF POWER DISTRIBUTIONS

SEED = 21212		SEED = 21212		SEED = 21212	
LL=1 UL=5		LL=1 UL=100		LL=1 UL=1000	
EXP = 2.7		EXP = -2.7		EXP = -2.7	
EVENTS	EXP	EVENTS	EXP	EVENTS	EXP
10	-2.1	10	-2.7	10	-3.0
20	2.9	20	-3.0	20	-3.1
30	-2.9	30	-2.9	30	-2.9
40	2.7	50	2.8	50	2.7
50	2.8	100	2.8	100	2.6
100	-2.7	200	-2.9	200	2.6
		400	2.9	500	2.7

TABLE 5.2

RESULTS OF MAXIMUM LIKELIHOOD TREATMENT OF BINNING ERRORS

EVENTS	LL	UL	BINS	EXP(GIVEN)	EXP(FIT)
30	1	2	4	2.7	3.1
50	1	3	5	-2.7	-3.2
100	1	3	10	-2.7	-3.15
200	1	3	20	-2.7	-2.87
500	1	5	20	2.7	2.82
500	1	5	40	-2.7	-2.81
500	1	10	20	-2.7	-2.79
100	1	20	40	-2.7	2.52

AS A FIRST EXERCISE WE WILL LOOK AT THE EFFECT OF A MEASURING UNCERTAINTY WHICH IS PROPORTIONAL TO ENERGY, NAMELY,

E(F) = E(I)*(1 + F*(0.5-RND))   ,   6.1

WHERE E(I) IS THE INCOMING ENERGY, E(F) IS THE READING FROM THE INSTRUMENT, F REPRESENTS A RELATIVE SPREAD AND RND IS A RANDOM NUMBER UNIFORMLY DISTRIBUTED BETWEEN ZERO AND ONE. FOR EXAMPLE, IF F = 0.1 THEN 0.95*E(I)   E(F)   1.05*E(I). TABLE 6.1 SHOWS THE EFFECT OF INTRODUCING A RANDOM ERROR SUCH AS INDICATED IN EQ 6.1. IT IS SEEN THAT A RATHER LARGE ERROR OF THIS TYPE CAN BE TOLERATED WITHOUT SERIOUS DEGRADATION OF OUR ABILITY TO EXTRACT A SPECTRUM.

TO ILLUSTRATE THE SECOND KIND OF ERROR, WE IMAGINE THAT THE INCOMING ENERGY E IS DEGRADED TO A READING E' ACCORDING TO THE DISTRIBUTION,

P(E,E') = A*(E (-B))*A'*EXP(-(E-E')/(h*E)   ,   6.2

WHERE A*E(-B) REPRESENTS THE INCOMING SPECTRUM AND THE REMAINING FACTOR REPRESENTS THE DEGRADATION OF THE INSTRUMENT. THE PRODUCT h*E IS THE AVERAGE DEGRADATION E-E'. THE RESULT OF USING THIS DISTRIBUTION IS SHOWN IN TABLE 6.2. IT IS SEEN THAT THIS TYPE OF ERROR MAKES THE SPECTRUM LOOK STEEPER BY TRANSFERRING EVENTS FROM HIGHER TO LOWER ENERGIES.

7. COSMIC RAY DATA FROM CERENKOV COUNTERS

WHEN A HIGH ENERGY NUCLEUS WITH SPEED V AND ATOMIC NUMBER Z TRAVERSES A REFRACTIVE MEDIUM WITH REFRACTIVE INDEX N, A LIGHT PULSE IS GENERATED ACCORDING TO THE FORMULA,

L = K*(Z 2)*(1 - (C/(V*N)) 2)   ,   7.1

WHERE C IS THE SPEED OF LIGHT IN VACUUM AND K IS A CONSTANT. IT IS SEEN FROM EQ 7.1 THAT L(MAX), THE MAXIMUM LIGHT PULSE AVAILABLE WHEN V/C = 1, IS GIVEN AS

L(MAX) = K*(Z 2)/(1   1/N 2)   .   7.2

THE INDEX OF REFRACTION N OF THE GAS IN THE COUNTER WAS 1.00115 SO THAT L(MAX)/(K*(Z 2)) = 435.53 .

THE CERENKOV COUNTER ACTS TO CONVERT THE ENERGY SPECTRUM,

N(E) = A * E^(-B)   ,   7.3

TO THE PULSE HEIGHT SPECTRUM P(L) WHERE

N(E) dE = P(L) dL   ,   7.4

TABLE 6.1

EFFECT OF SIMPLE ENERGY UNCERTAINTY ON SPECTRUM EXTRACTION *

P(E) = A E (-B)   B = 2.7   LL = 1   UL =5

E(F) = E(I)*(1 + h*(0.5 - RND))

EVENTS	ERROR PARAMETER h	EXTRACTED B
10	0.1	2.62
10	0.2	2.59
10	0.3	2.57
10	0.8	2.57
10	0.9	2.63
10	1.9	3.84

TABLE 6.2

EFFECT OF EXPONENTIAL DEGRADATION ON SPECTRUM EXTRACTION

P(E) = A E (-B)   B = 2.7   LL = 1   UL = 5

P(E,E') = A*E (-B)*A'EXP [-(E-E')/(h*E)]

EVENTS	ERROR PARAMETER h	EXTRACTED B
10	0.01	2.71
10	0.10	3.04
10	0.20	4.53
10	0.30	9.90
10	0.50	****
30	0.001	3.00
30	0.01	3.09
30	0.05	3.52
30	0.10	4.36
30	0.20	****

**** PROGRAM FAILED TO RUN

OR

P(L) = N(E) (dE/dL) . 7.5

THE KINETIC ENERGY E OF A PARTICLE OF MASS M IS GIVEN BY THE EXPRESSION,

E = (GAMMA -1)*M*C 2 , 7.6

WHERE

GAMMA = SQR(1 / (1-BETA 2)) ; BETA = V/C , 7.7

AND C IS THE SPEED OF LIGHT. IT IS CONVENIENT TO EXPRESS ENERGY IN UNITS OF MC 2 SO THAT E = GAMMA-1. AFTER A LITTLE ALGEBRA WE FIND THAT

E = [X/(X-1)] - 1 , 7.8

WHERE

X = (N 2)*(1 - (L/K)) . 7.9

USING THESE DEFINITIONS,

dE/dL = [(N 2)/(2*K)]*[X*(X-1) 3] (-1/2) 7.10

THE NORMALIZATION CONSTANT A IN EQ 7.3 IS FOUND BY DIRECT INTEGRATION OF EQ 7.3 TO BE

A = (B-1)/[LL (B+1) - UL (-B+1)] , 7.11

WHERE LL IS THE LOWER ENERGY LIMIT CONSIDERED FOR THE DISTRIBUTION AND UL IS THE UPPER LIMIT. PUTTING EQS 7.3, 7.8, 7.10, AND 7.11 TOGETHER, WE FIND AFTER A LITTLE ALGEBRA THAT

P(L) = [(A*N 2)/(2*K)]*[1/(SQR(X)-SQR(1-X))] B *
SQR(1/X) * (X-1) ((B-3)/2) 7.12

THIS DISTRIBUTION IN PULSE HEIGHT IS RATHER FLAT WHEN B=2.9, IS MONOTONIC UPWARD FOR B < 2.9 AND IS MONATONIC DOWNWARD FOR B 2.9. THE (B-3) EXPONENT IN THE LAST FACTOR OF EQ 7.12 IS LARGELY RESPONSIBLE FOR THIS BEHAVIOIR. AN IMPORTANT FEATURE OF THESE CURVES IS THAT, AS B DECREASES FROM 3, THE STACKING OF THE DATA POINTS JUST BELOW L(MAX) BECOMES MORE PRONOUNCED. FOR B 3 , P(L) IS RATHER EVENLY MONATONIC DOWNWARD FROM BEGINNING TO END.

THE PROBABILITY OF ONE COUNT WITH L(I) < L < L(I+1) IS P(L(I)) SO THAT THE PROBABILITY OF N(I) COUNTS IN THE INTERVAL IS P(L(I)) N(I). A GOOD MAXIMUM LIKELIHOOD FUNCTION P IS GIVEN AS

P = PRODUCT(I) P(L(I)) N(I) , 7.13

OR

LOG P = SUM(I) N(I) * LOG[P(L(I))] . 7.14

THIS FUNCTION IS EASILY MAXIMIZED BY THE METHODS DEVELOPED IN THE ATTACHED COMPUTER PROGRAMS. IF WE DO NOT BIN THE DATA, A CONVENIENT MAXIMUM LIKELIHOOD FUNCTION IS GIVEN AS

LOG P = SUM(I) LOG P[L(I)] 7.15

IF WE LET B AND L(MAX) BE VARIABLES, WHERE L(MAX) MUST BE GREATER THAN ANY PULSE HEIGHT L(I) USED, WE CAN FIND B AND L(MAX) WHICH MAXIMIZE LOG P FOR THE DATA. ANY ANSWER 2 B 4 IS POSSIBLE FOR SMALL VARIATIONS IN L(MAX). THE BEST AND MOST BELIEVABLE RESULTS WERE OBTAINED BY MAXIMIZING [LOG(P)]/N WITH B AND L(MAX) AS VARIABLES, WHERE N IS THE TOTAL NUMBER OF PARTICLES USED IN THE MAXIMIZATION. IT SHOULD BE NOTED THAT P(L) IS RATHER FLAT AND THAT ANY FUNCTION WHICH RANDOMLY CHANGES THE SIZE OF THE LIGHT PULSES WILL NOT CHANGE TO ANY APPRECIABLE EXTENT THE SHAPE OF P(L) PREDICTED BY EQUATION 7.12, EXCEPT AT THE ENDS. DATA WITH L(I) L(MAX) IS REFLECTED THROUGH L(MAX) TO A LOWER ENERGY 2*L(MAX) L(I). THE ARGUMENT IN FAVOR OF THIS IS AS FOLLOWS. IF WE IMPOSE TO THE RIGHT OF P(L) A MIRROR IMAGE P'(L), THE PROBABILITIES EXCHANGED BETWEEN THE TWO DISTRIBUTIONS ARE LIKEWISE MIRROR IMAGES. SINCE EXCHANGES ARE EQUAL, THE LEAKAGES AT THE BOUNDARIES ARE SUFFICIENT TO PRESERVE P(L) IF THEY ARE REFLECTED BACK IN.

THE RESULTS OF TREATING CARBON AND IRON COSMIC RAYS BY THE ABOVE MAXIMUM LIKELIHOOD TREATMENT ARE SHOWN IN FIG 7.1 AND TABLE 7.1.

8 SUMMARY

SEVERAL USEFUL TECHNIQUES WERE DEVELOPED TO EXTRACT PARAMETERS FROM DATA ON RANDOM EVENTS. THE BEST OF THEM WERE BASED ON MAXIMUM LIKELIHOOD STATEMENTS. THESE METHODS PROVED SUCCESSFUL IN TREATING DATA FROM COSMIC RAYS FOR THE CASES OF CARBON AND IRON NUCLEI.

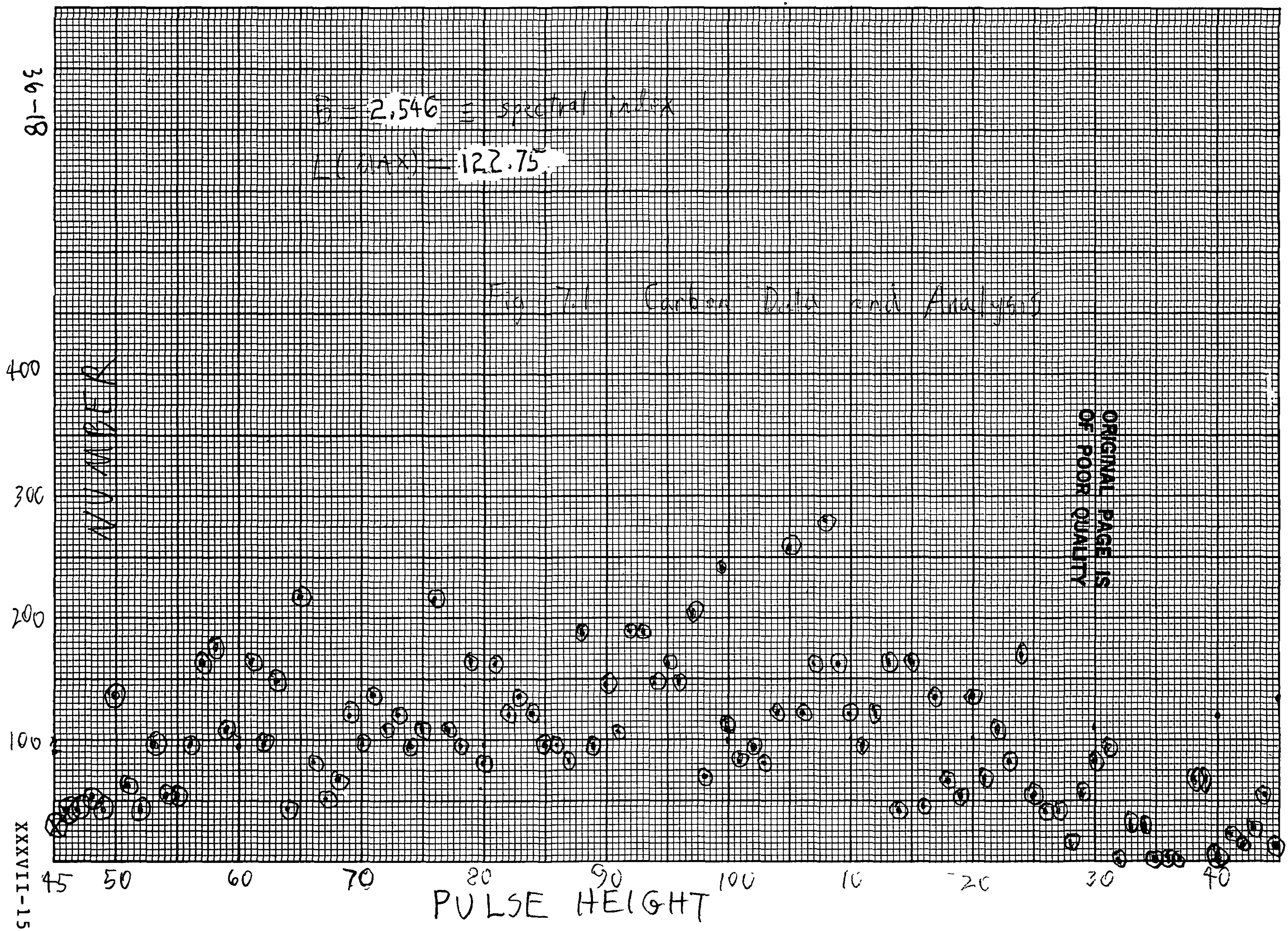

Fig 7.1 Carbon Data and Analysis

TABLE 7.1

ANALYSIS OF DATA ON IRON NUCLEI

[illegible]

NUMBER OF DATA POINTS 82

STARTING SPECTRAL INDEX FOR THE MAXIMUM LIKELIHOOD ITERATION 2.7

STARTING L(MAX) 2114

FINAL SPECTRAL INDEX 2.0938

FINAL L(MAX) 2022.5

```
10 ' PROGRAM TO FIT P(L) TO IRON DATA UL=L(MAX)-1 L(I) > L(MAX) REFLECTED
20 ' B:NASA48D.BAS   R. D. SHELTON 29 JULY 85 09:32
30 CLS
40 DIM L(200),LL(200),P#(200),E#(200)
50 DATA 1230,767,1937,1950,1986,1628,2466,1367,2079,2243
60 DATA 1974,1895,2107,1968,2195,1965,468,2023,1602,1958
70 DATA 632,1217,1018,2034,1494,911,1207,1948,1880,1947
80 DATA 1828,2003,944,901,1990,1965,1805,1070,808,1463
90 DATA 1647,1370,625,640,1419,484,1951,662,613
100 DATA 1663,1264,1330,2103,1726,1125,2324,630,2468,1133
110 DATA 642,1822,2176,1959,1154.754,1841,1738,2259,977
120 DATA 1627,1538,1226,1730,2034,1874,2438,1448,812,2242
130 DATA 1570,852,2069
140 SUM#=0
150 P$="#####.####" :PP$="### "
160 FOR I=1 TO 82
170      READ LL(I)
180      ' SUM#=SUM#+LL#(I)
190      ' PRINT I,LL(I):IF I MOD 20 = 0 THEN INPUT Z$
200 NEXT I
210 FOR I=1 TO 82
220     FOR J=1 TO 82
230             IF LL(I) LL(J) THEN SWAP LL(I),LL(J)
240     NEXT J
250     IF I MOD 10 =0 THEN PRINT I;
260 NEXT I
270 ' FOR I=1 TO 82
280 '      LPRINT I;"  ";LL(I)
290 ' NEXT I
300      ' PREPARE FOR ITERATION
310 CLS:PRINT
320 PRINT
330    HM=1:HB=1:IT=1:PRINT:INPUT " ENTER STARTING EXPONENT    ",M#
340 PRINT
350    INPUT " ENTER L(MAX)    ",B#
370    ET#=1.00115:ET2#=ET# 2
380    PRINT:INPUT " ENTER STEP FOR EXPONENT  ",Z$
390    IF Z$  "" THEN HM#=VAL(Z$)
400    PRINT:INPUT " ENTER STEP FOR L(MAX)  ",Z$:PRINT
410    IF Z$ <"" THEN HB#=VAL(Z$)
420    IF IT 1 THEN GOTO 460
430   'GOSUB 610
440    S1#=S#
450    'ITERATION SCHEME
460       GOSUB 810
470       GOSUB 940
480       LOOP=IT MOD 10:PRINT
490       IF LOOP< 0 THEN GOTO 460 ELSE GOTO 500
500 PRINT
510 INPUT "CHOICE ? CR-GO ;C-CHANGE STEP ;P-LPRINT LL(I),L(I);QQ-QUIT ",Z$
520       PRINT
530       IF Z$="QQ" THEN PRINT I;" ";IT;" EXP =  ";M#;"   L(MAX) =   ";B#
540       IF Z$="C" THEN GOTO 380
550       IF Z$="P" THEN GOTO 970
560       IF Z$="QQ" THEN GOTO 600
580       IF Z$="QQ" THEN INPUT Z$
590    GOTO 460
600 END
610  'SBR TO COMPUTE SUMS FOR ITERATION
620 S#=0:LL=468:K#=B#/(1-1/ET2#)
630 FOR JJ=1 TO 82
640     IF LL(JJ)<=B#-.5 THEN L(JJ)=LL(JJ) ELSE L(JJ)=2*B#-LL(JJ)
650 NEXT JJ
660 X#=ET2#*(1-LL/K#):SLL#=SQR(X#/(X#-1))-1
670 X#=ET2#*(1-(B#-.5)/K#):SUL#=SQR(X#/(X#-1))-1
```

ORIGINAL PAGE IS OF POOR QUALITY

```
680 [illegible]
690    PRINT [illegible] INPUT Z[illegible]
700    PRINT [illegible]
710 FOR J=1 TO [illegible]
720       IF [illegible] THEN X#=ETL#(J-L(JJ)/[illegible]) ELSE [illegible]=ETL#-(1 [illegible])
730       E#=INT(X#/(Z# 1)) 1
740       [illegible]
750       [illegible]
760       IF [illegible] INT [illegible]
[illegible]
770       [illegible]
780          PRINT USING [illegible]
790 NEXT [illegible]
800 [illegible]
810 [illegible] DETAIL [illegible]
820 [illegible]
830 GOSUB [illegible]
840 IF [illegible] THEN [illegible]
850 [illegible] THEN [illegible]
860 [illegible]
870 [illegible]
880 [illegible] THEN [illegible]
890 [illegible] THEN [illegible]
900 [illegible]
910 [illegible]
920 [illegible]
930 [illegible]
940 [illegible] FOR [illegible]
950 PRINT USING [illegible]
[illegible]
960 [illegible]
970 FOR I=1 TO [illegible]
980 PRINT [illegible]
990 NEXT I
1000 [illegible]
```

ORIGINAL PAGE IS OF POOR QUALITY