SOFTWARE MODELS:

A BAYESIAN APPROACH TO PARAMETER ESTIMATION IN THE JELINSKI-MORANDA SOFTWARE RELIABILITY MODEL

by

Bev Littlewood, The City University, London, England
Ariela Sofer, The George Washington University, Washington, D.C.

Abstract

Maximum likelihood estimation procedures for the Jelinski-Moranda software reliability model often give misleading answers. We show here that a reparameterization and a Bayesian analysis eliminate some of the problems incurred by MLE methods and often give better predictions on sets of real and simulated data.

Practical difficulties in estimating the initial number of errors N and the failure rate of each error ϕ by the method of maximum likelihood are:

1. $\hat{N}$, the MLE of N, is occasionally infinite (i.e., the routines for calculating $\hat{N}$ and $\hat{\phi}$ do not converge). Littlewood and Verrall show that $\hat{N}$ is finite if and only if the regression line of the interevent times t_i vs. i has positive slope.
2. A serious problem is that often $\hat{N} \simeq n$, the sample size, and sometimes $\hat{N} = n$. Thus the MLE predicts that the program is perfect even when it is far from being so. Forman and Singpurwalla have shown that $\hat{N}$ and $\hat{\phi}$ can only be trusted near the end of debugging, i.e., when almost all failures have been removed.

3. Even when these problems are not encountered, the results obtained from the model are too optimistic; it predicts the reliability to be greater than it really is.

In view of these deficiencies, we are led to consider a Bayesian approach to the estimation problem. It seems plausible that it is easier to correctly estimate the initial program failure rate $\lambda = N\phi$ than the initial number of bugs N, since small errors in $\hat{\phi}$ could lead to large errors in $\hat{N}$. It is therefore plausible to reparameterize the model to (λ,ϕ) instead of (N,ϕ).

Using now the Bayesian approach, letting prior (λ,ϕ) = prior $(\lambda)\cdot$ prior (ϕ), where prior (λ) and prior (ϕ) are gamma distributed, and using

$$R_{n+1}(t) = P(T_{n+1} < t) = P(T_{n+1} < t \mid t_1,\ldots,t_n)$$
$$= \int P(T_{n+1} < t \mid \lambda,\phi)\, \mathrm{post}(\lambda,\phi \mid t_1,\ldots,t_n)\, d\lambda d\phi$$

we obtain an explicit estimate of the program's current reliability. Similarly, we can get in closed form the distributions of the number of bugs remaining in the program, the number of bugs that have to be removed in order to attain a given reliability, and the times between future consecutive failures (provided they are well defined, i.e., the program is not perfect).

The quality of these estimations was examined for the special case when λ and ϕ have an (improper) uniform prior distribution over $[0,\infty)$ (i.e., a noninformative prior distribution). The predictions were examined both for real and for simulated sets of data. In all cases where ML erroneously predicts the program to be perfect, the Bayesian method gives a positive probability that the program is not perfect. Moreover, since the predicted reliability is given in closed form, problems of convergence of the computer program are not encountered.

To examine the quality of prediction, we use a goodness of fit procedure. Suppose that from the data $t_1,\ldots,t_n$ we predict the distribution of T_{n+1}, the time to next failure. We then observe t_{n+1}. Define $U_n = \Pr(T_{n+1} < t_{n+1})$. If the model is correct, then U_n are uniform variables on $(0,\ell)$. We compare the sample c.d.f. of the u_n's with a line of unit slope which is the uniform c.d.f.

When applying the goodness of fit procedure to real data sets, the Bayesian approach is almost always better than the MLE method. For the simulated data, the goodness of fit procedure on the Bayesian estimates give very good results; this, however, is not always true for the real data sets.

There seems to be evidence that the J-M model is intrinsically optimistic in its estimate of software reliability. This could be a consequence of the assumption that all errors contribute equally to the failure rate. A new model by Littlewood relaxes this assumption with the result that earlier fixes tend to involve larger reductions in the failure rate than the later ones. It can be shown that this model is less optimistic than the J-M model and we hope to examine its performance on real and simulated data in future work.

THE VIEWGRAPH MATERIALS
for the
B. LITTLEWOOD/A. SOFER PRESENTATION FOLLOW

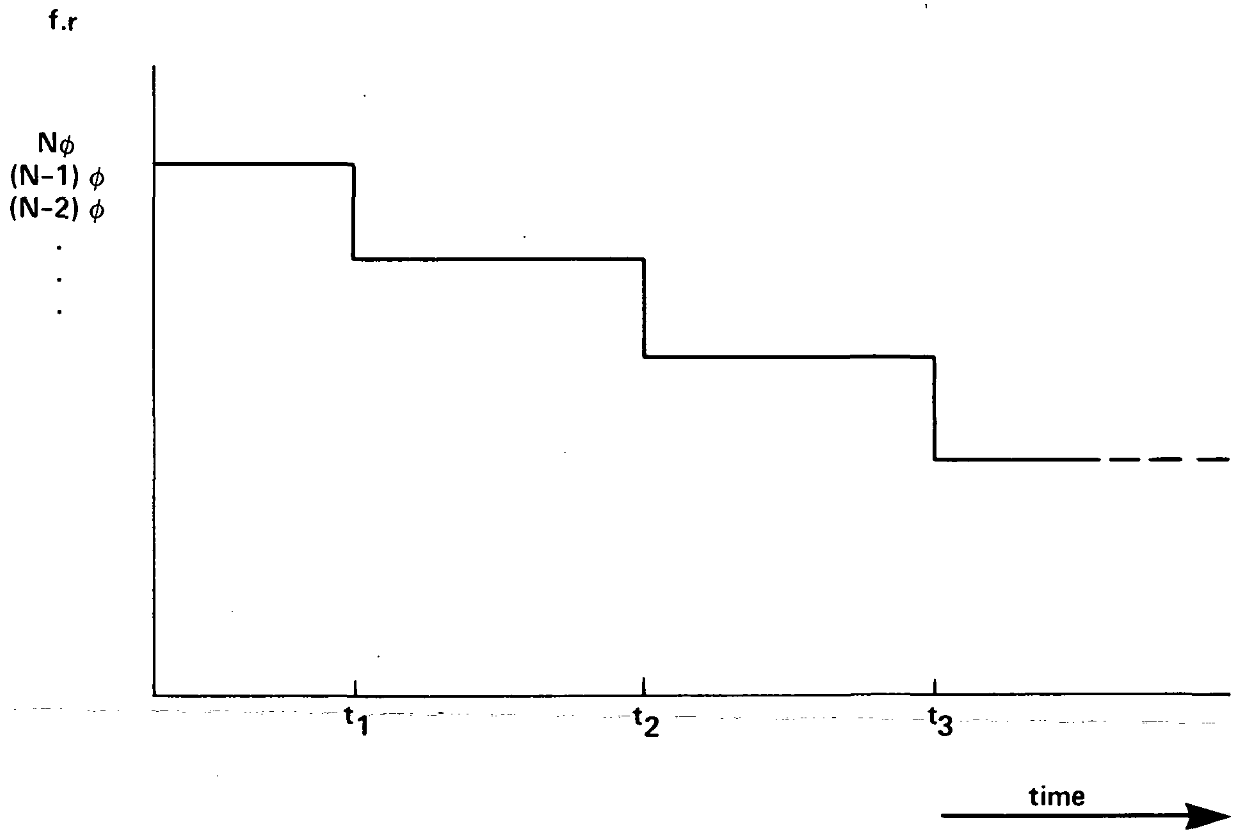
JELINSKI-MORANDA model assumptions:

1. Successive inter-failure times $T_1, T_2, \ldots$ are independent.

$$\text{pdf}(t_i \mid \lambda_i) = \lambda_i e^{-\lambda_i t_i}$$

2. $\lambda_i = (N - i + 1)\,\phi$ where

   N is "initial number of faults" ϕ is "contribution to program failure rate from each fault"



Note that
1. All fixes have same effect.
2. Same model by SHOOMAN and MUSA. Same assumptions for NHPP model by GOEL-OKUMOTO.

There seems to be 3 problems with J-M:

1. $\hat{N}$ occassionally infinite ($\hat{\phi} = 0$)

   Nec. & Suff. conditions: "Regression line of t_i versus i has negative slope"
   (Littlewood, Verrall: 1981 IEEETR)

   This can also occur with simulated data from J-M with finite N, $\phi \neq 0$,

   However $\hat{\lambda} = N\hat{\phi}$ is finite, non-zero.

2. Reliability predictions always(?) too optimistic

3. $\hat{N}$ usually too small, sometimes equal to sample size (i.e. program is "perfect")

Table 7.
Failure Intervals – System 3 System Test Phase

i	T_i	
1	115,	1
2	0,	1
3	83,	3
4	178,	3
5	194,	3
6	136,	3
7	1077,	3
8	15,	3
9	15,	3
10	92,	3
11	50,	3
12	71,	3
13	606,	6
14	1189,	8
15	40,	8
16	788,	18
17	222,	18
18	72,	18
19	615,	18
20	589,	26
21	15,	26
22	390,	26
23	1863,	27
24	1337,	30
25	4508,	36
26	834,	38
27	3400,	40
28	6,	40
29	4561,	42
30	3186,	44
31	10571,	47
32	563,	47
33	2770,	47
34	652,	48
35	5593,	50
36	11696,	54
37	6724,	54
38	2546,	55
39	−10175,	56

SYSTEM 3	$\hat{N}$		$\hat{\phi}$
FAILURE NUMBER	ESTIMATED FAILURES	ESTIMATED INITIAL MTTF	MORANDA PHI
2	999999	0.5750E+02	0.173913E-07
3	999999	0.6600E+02	0.151515E-07
4	5	0.5900E+02	0.338983E-02
5	6	0.6480E+02	0.257202E-02
6	8	0.7275E+02	0.171821E-02
7	7	0.7884E+02	0.181206E-02
8	8	0.8845E+02	0.141318E-02
9	12	0.1196E+03	0.696972E-03
10	19	0.1396E+03	0.377017E-03
11	55	0.1609E+03	0.112990E-03
12	999999	0.1688E+03	0.592304E-08
13	22	0.1387E+03	0.327621E-03
14	15	0.1125E+03	0.592367E-03
15	18	0.1306E+03	0.425447E-03
16	18	0.1306E+03	0.425294E-03
17	21	0.1476E+03	0.322715E-03
18	25	0.1616E+03	0.247463E-03
19	25	0.1622E+03	0.246615E-03
20	25	0.1612E+03	0.248210E-03
21	31	0.1807E+03	0.178535E-03
22	33	0.1854E+03	0.163413E-03
23	26	0.1609E+03	0.239046E-03
24	26	0.1606E+03	0.239457E-03
25	25	0.1520E+03	0.263205E-03
26	26	0.1628E+03	0.236199E-03
27	27	0.1764E+03	0.210001E-03
28	28	0.1876E+03	0.190384E-03
29	29	0.2023E+03	0.170456E-03
30	30	0.2182E+03	0.152766E-03
31	31	0.2427E+03	0.132935E-03
32	32	0.2642E+03	0.118265E-03
33	33	0.2853E+03	0.106202E-03
34	34	0.3041E+03	0.967196E-04
35	35	0.3248E+03	0.879556E-04
36	36	0.3519E+03	0.789439E-04
37	37	0.3804E+03	0.710397E-04
38	38	0.4073E+03	0.646041E-04

How well does model perform?

Simplest problem is estimation of current reliability:

Given data $t_1, \ldots, t_{i-1}$, what can we say about T_i?

What is cdf $F_i(t)$?

Obtain ML estimates of N, ϕ, based on $t_1, \ldots, t_{L-1}$ and use "Predictor distribution"

$$\hat{F}_i(t) = 1 - e^{-(\hat{N}-i+1)\hat{\phi}t}$$

If prediction is "good"

$U_i = \hat{F}_i(T_i)$ is approx.

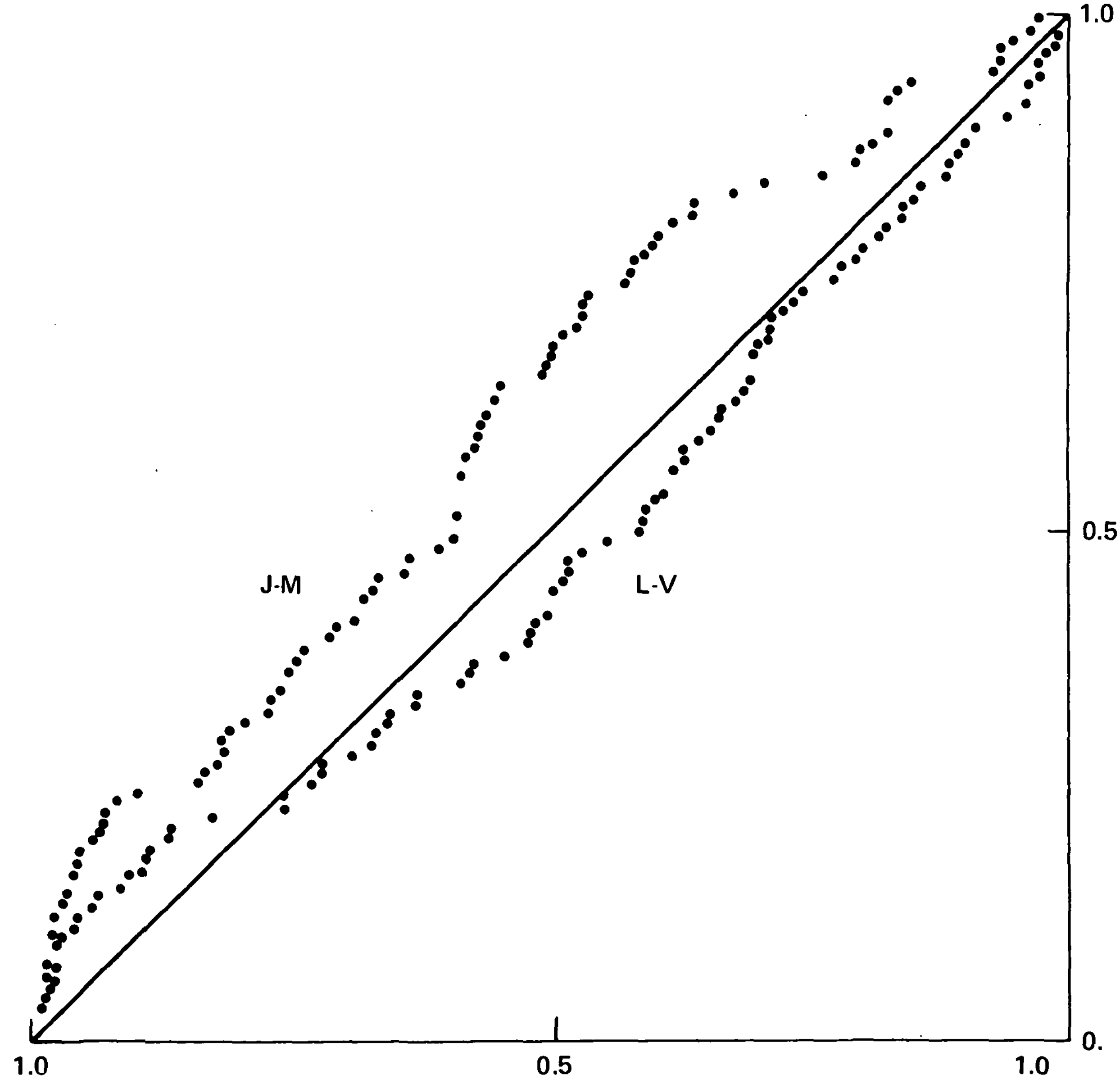
U(0,1). Examine Q-Q plots of realizations U_i

EXAMPLE

Data: MUSA "System 1", range of i:30-129

Jelinski-Moranda: poor prediction, *optimistic*

Littlewood-Verrall: good prediction, slight pessimism



Bayesian J-M

Reparameterize to (λ, ϕ) from (N, ϕ) where $\lambda \equiv N\phi$ "initial failure rate".

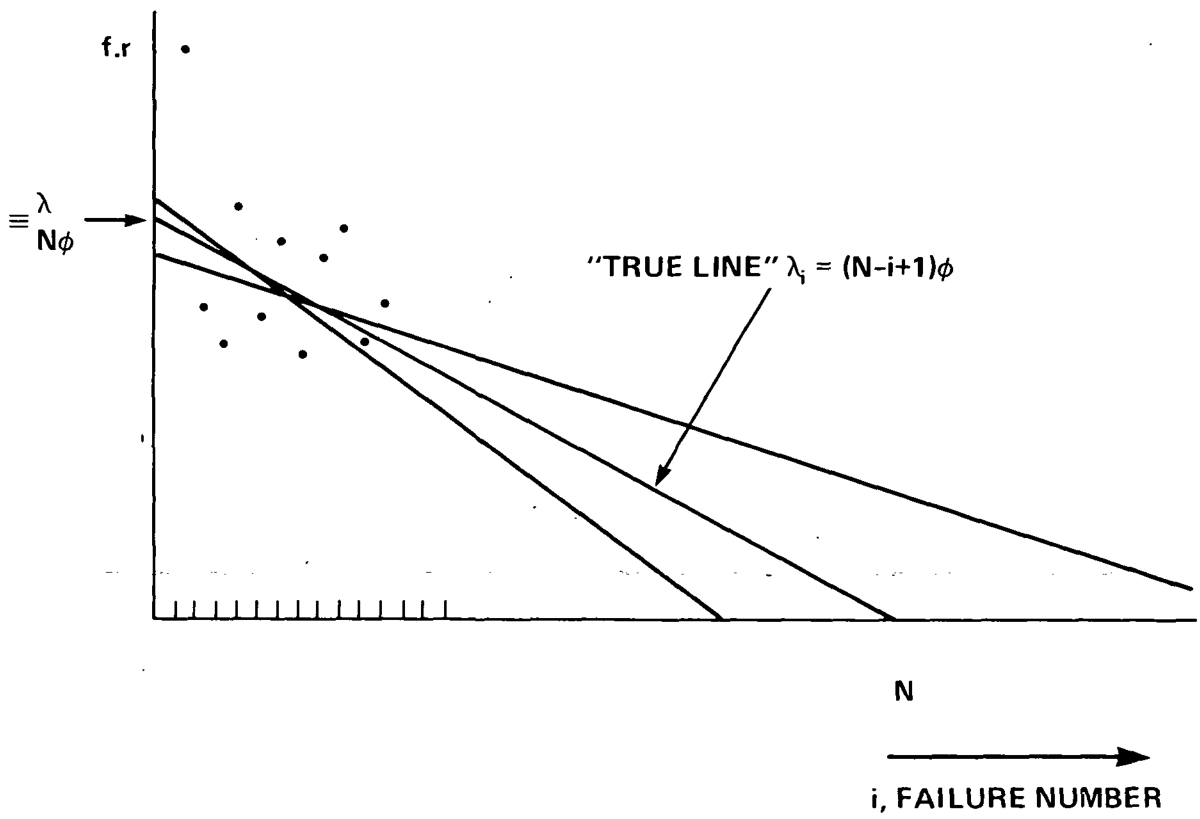
Assume:

prior (λ, ϕ) = prior $(\lambda) \cdot$ prior (ϕ) where prior (λ) and prior (ϕ) are gamma distributed

Then "predictor distribution" is

$$\tilde{F}_1(t) \equiv P(T_i < t) \equiv P(T_i < t \mid t_1, \ldots t_{i-1}) = \int P(T_i < t \mid \lambda, \phi) \text{ post } (\lambda, \phi \mid t_1 \ldots t_{i-1}) \, d\lambda d\phi$$

Reparameterization: Informal Justification



For the case of uniform (improper) priors we get:

$$F_{i+1}\,(t \mid t_i. .t_i) =$$

$$c\left[\sum_{k=0}^{i} \frac{a_{k,i}\,k!\,(i-k)!}{\left(\sum_{j=1}^{i}(i-j+1)\,t_j\right)^{k+1}}\left(\frac{1}{\left(\sum_{1}^{i} t_j\right)^{i-k+1}} - \frac{1}{\left(t+\sum_{1}^{i} t_j\right)^{i-k+1}}\right)\right]$$

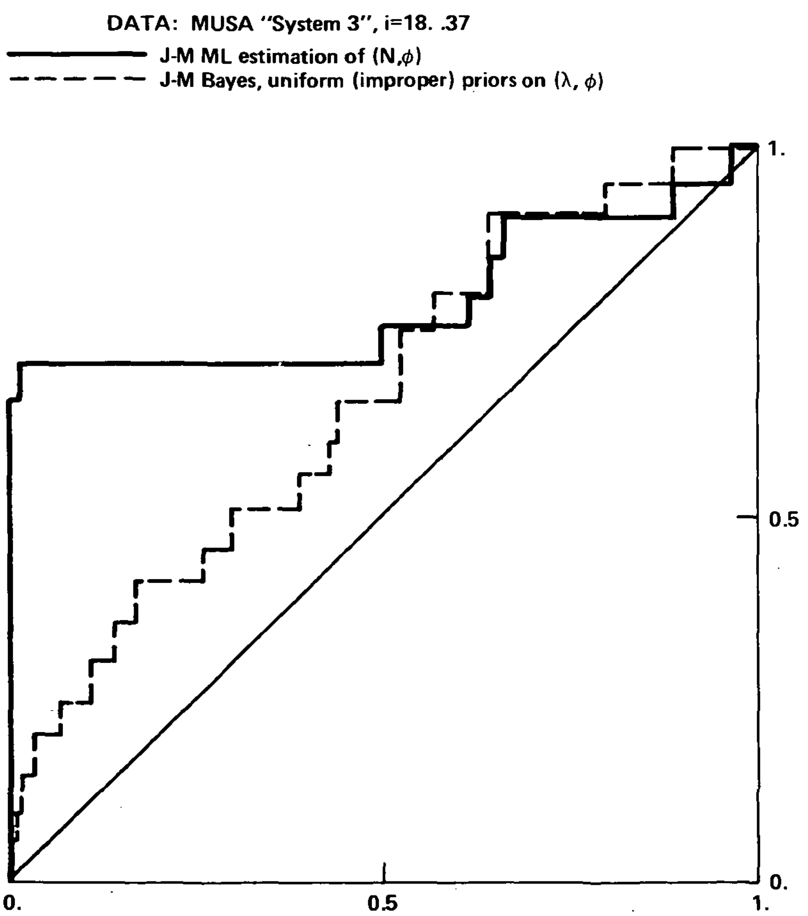
where $c^{-1} = \sum_{k=0}^{i-1} \dfrac{a_{k,\,i-1}\,k!\,(i-k)!}{\left(\sum_{j=1}^{i}(i-j)\,t_j\right)^{k+1}\left(\sum_{j=1}^{i} t_j\right)^{i-k+1}}$

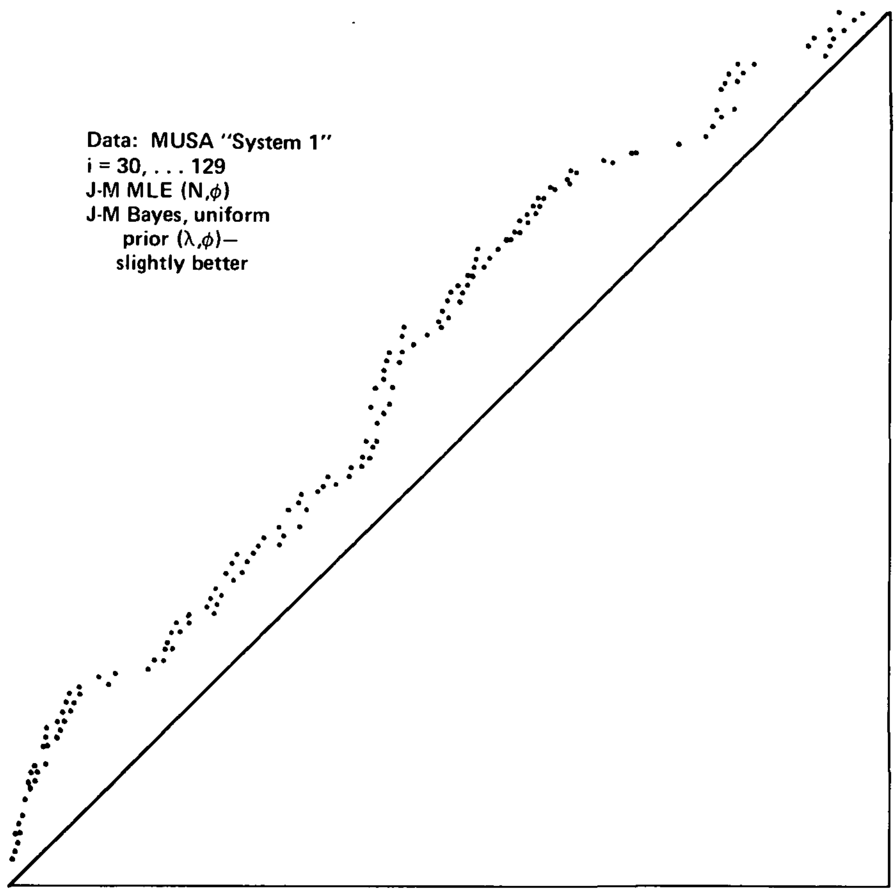
and where $a_{k,i}$ is the coefficient of x^{i-k} in $\prod_{k=1}^{i}(x+k) = \sum_{k=0}^{i} a_{k,i}\,x^{i-k}$

These coefficients are easily computed from the relation

$$a_{k,i} = i\,a_{k-1,\,i-1} + a_{k,i-1} \quad k \geqslant 1$$

$$a_{01} = 1 \quad a_{11} = 1 \quad a_{0i} = 1 \quad \forall_i$$





Conclusion!

1. Bayes J-M seems always (?) better than MLE J-M, but sometimes only slightly.

2. Results on *real* data are always optimistic.

3. But on SIMULATED data from J-M model, Bayes is very good, ML poor

⇒ real data do not follow J-M model?

Hypothesis: Assumption of equal ϕ's is wrong. In fact ϕ's different. Larger ones tend to be eliminated earlier: