DOES AN OBSERVED SEQUENCE OF NUMBERS FOLLOW A SIMPLE RULE?

(Another Look at Bode's Law)

BY

BRADLEY EFRON

TECHNICAL REPORT NO. 8

NOVEMBER 26, 1969

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT GP-15909

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

DOES AN OBSERVED SEQUENCE OF NUMBERS FOLLOW A SIMPLE RULE?

(Another Look at Bode's Law)

By

Bradley Efron

TECHNICAL REPORT NO. 8

NOVEMBER 26, 1969

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT GP-15909

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Does an Observed Sequence of Numbers Follow a Simple Rule?
(Another Look at Bode's Law)

By

Bradley Efron

0. Summary.

Bode's law describes the distances of the planets to the sun by the simple formula $a+b\cdot 2^n$, where a and b are constants and n is the order from the sun (so Mercury is represented by $n=1$, Venus by $n=2$, etc.). The actual planetary distances do not follow this description perfectly, and the question arises whether Bode's law is real or artifactual. In a recent paper I. J. Good has used Bayesian methods to conclude that there is strong evidence in favor of the validity of Bode's law. Here we reanalyze the observed planetary distances using non-Bayesian significance testing procedures, and reach the opposite conclusion.

The three purposes of this paper are (1) to reexamine the validity of Bode's law; (2) to develop some insight into the more general problem of testing whether or not an observed sequence of numbers follows some simple rule; and (3) to discuss the philosophical basis of Fisherian significance testing. It is not intended to provide an example of disagreement between Bayesian and non-Bayesian methods. Rather the two analyses are seen to be similar in execution, the different conclusions arising from different choices of the statistical models.

A method previously suggested by Good for relating Bayesian to non-Bayesian hypothesis tests is shown to be valid in the context of this problem.

1. Introduction.

Often a sequence of numbers will appear to follow some simple "law", even though no theoretical explanation for the law can be found. A classic example is Bode's law, which describes the mean distances from the sun to each of the planets by the simple geometric progression

$$b_n = a + b \cdot 2^n, \quad n = 1, 2, 3 \ldots .$$

Here n is the order from the sun, (so Mercury is represented by $n=1$, Venus by $n=2$, etc.) and a and b are constants chosen to give a good fit to the actual distances, subject to the constraint that b_3, the Earth's distance, equals 1. These constants are

$$a = .4, \quad b = .075 .$$

Table 1 compares the actual distances, d_n, with the Bode numbers:

TABLE 1. Bode Distances and Observed Distances

Planet	n	Bode Distance b_n	Actual Distance d_n
Mercury	1	0.55	0.387
Venus	2	0.70	0.723
Earth	3	1.00	1.000
Mars	4	1.60	1.524
Asteroids	5	2.80	2.9
Jupiter	6	5.20	5.203
Saturn	7	10.00	9.546
Uranus	8	19.60	19.20
Neptune	9	38.80	30.09
Pluto	10	77.20	39.5

The obvious question is whether Bode's law is "real", or whether it is simply an ingenious numerical artifact of Bode's imagination. We can make this a statistical question by specifying

(i) a statistical model* of what we mean by Bode's law being real

and

(ii) an alternative, less interesting, statistical model describing the situation where Bode's law is an artifact.

Once the two models have been agreed upon, the question of the validity of Bode's law reduces to a problem of hypothesis testing.

I. J. Good, in a provacative recent paper [3], has carried out this program**. Using Bayesian methods to reduce his models (i) and (ii) to simple hypotheses (in the statistical sense that each hypothesis is reduced to a single probability distribution), he concludes that there is a likelihood ratio of between 300:1 and 700:1 in favor of Bode's law being real rather than artifactual. In Section 3 I will present another analysis of the planetary distances that is quite a bit less favorable to Bode's law. Although my analysis will be carried out in a non-Bayesian framework this is not intended to be an example of disagreement between

---

*A "statistical model" is a family of probability distributions.

**I am using Good's conventions and notations in this note. In particular, the value $n=1$ for Mercury is his emendation of Bode's original law, which used $n=-\infty$. The reader is referred to Good's paper for an extensive discussion and bibliography of Bode's law.

objective and subjective statistical methods. Actually the two analyses are rather similar in execution, the real difference lying in the choice of statistical models for (i) and (ii).

2. A Critique of Good's Models.

Good proposes the following statistical models for what he calls "B", the hypothesis that Bode's law is true, and "$\bar{B}$", the hypothesis that it is not true:

(i) Under B the logarithms of the planetary distances for $n=2,3,\ldots,8$ are independently normal*,

$$\log D_n \sim \mathcal{N}(\log(a + b \cdot 2^n), \sigma^2), \quad n = 2, 3, \ldots, 8 .$$

(ii) Under $\bar{B}$ the log distances $\log D_n$ for $n=2,3,\ldots,8$ are the order statistics of 7 independent uniform random variables on some interval $[\log d, \log D]$. In both cases Good applies Bode's law only to the planets Venus through Uranus, $n=2$ to 8, because the law obviously fails for $n=1$, $n=9$, and $n=10$. He reduces his final likelihood ratio for B to $\bar{B}$ by a factor of about 5 as a penalty for ignoring an unpleasant part of the data.

There are three nuisance parameters in model (i), a, b, and σ, and Good assigns these independent "log Cauchy" prior distributions, with medians at their respective maximum likelihood estimates. That is, letting $c(x)$ be a standard Cauchy density function,

*Following convention, D_n is the random variable whose observed value is d_n.

$$c(x) = \frac{1}{\pi} \frac{1}{1+x^2} ,$$

$\log a$ is assigned the prior density function $c(\log a - .4)$, $\log b$ is assigned $c(\log b - .075)$, and $\log \sigma$ is assigned $c(\log \sigma - .032)$, mutually independently. (The values .4, .075, and .032 are maximum likelihood estimates from model (i).)

There are two nuisance parameters in model (ii), $\log d$ and $\log D$, but Good reduces this to one by tacitly assuming that the interval $[\log d, \log D]$ has the same midpoint as the observed interval $[\log d_2, \log d_8]$, namely $\frac{1}{2} \log d_2 d_8$. He then assigns $\log \frac{D_8}{D_2}$ a log Cauchy distribution with median $\log \frac{d_8}{d_2}$, the maximum likelihood estimate from model (ii).

Integrating the likelihood function under model (i) for the observed data $d_2, d_3, \ldots, d_8$ over the prior densities of $a, b,$ and σ gives a marginal likelihood of 348 for the data under model (i). The marginal likelihood under model (ii) is 0.163, yielding a likelihood ratio of about 2000:1 in favor of B over $\bar{B}$. After assessing the penalty factor for ignoring $d_1, d_9,$ and d_{10}, Good arrives at a likelihood ratio of between 300:1 and 700:1 in favor of B over $\bar{B}$.

Three criticisms of Good's formulation are given below, in descending order of importance. To reiterate, these are criticisms of his statistical models and not of the Bayesian methodology. If one accepts Good's (i) and (ii) then Bayes rule is not at all essential to the conclusion that Bode's law is valid. For example the Neyman - Pearson likelihood ratio statistic for B versus $\bar{B}$ has a value of about 6000, which if accepted at face value would terminate our investigation.

Criticism 1. Unlikely choice of the alternative to Bode's law.

Quoting Good, page 30, "since it is obvious that $d_1, d_2, \ldots, d_{10}$ cannot be regarded as selected from a uniform distribution there is no point in contrasting Bode's law with that hypothesis: if we did so it would be obvious that Bode's law was not accidental". The point is one familiar to politicians: if you state your opponent's case absurdly enough, your own position will look good by comparison.

Unfortunately Good's uniform distribution for the log distances is also an unlikely alternative to Bode's law, given the general nature of the planetary distances, (though it is better than a uniform distribution on the raw distances). In particular notice that the differences $d_{n+1}-d_n$ are themselves increasing over the range of Good's interest, $n=2$ to 7, and that this "law of increasing differences" also applies to the next case, $n=8$, barely failing for the last difference $n=9$.

If one believes that the law of increasing differences is not accidental, but rather a reasonably dependable result of whatever mechanism determines planetary distances, then this fact should be incorporated into any hypothesis proposed as an alternative to the stronger hypothesis of Bode's law. This is not the case for the log uniform hypothesis. For example, a random division of $[\log d_2, \log d_8]$ into 6 intervals by the uniform choice of 5 interior points, when converted back to distance units by exponentiation, yields increasing differences only 3% of the time. (Computer simulation yielded 61 such cases in 2000 repetitions.)

Thus in Good's statistical formulation Bode's law will receive a great deal of credit for predicting increasing differences, which it very dependently does, as opposed to Good's null hypothesis which seldom does. This weakens Good's conclusion, since only a straw man has been eliminated from contention.

Criticism 2. Over-likely choice of the model for Bode's law being true. Suppose that the planetary distances approximately were in the proportions 1, 4, 9, 16, ... It seems certain that Bode or one of his colleagues would have noticed this and proposed a law in the form $a+b\cdot n^2$. Good seems to agree with this sentiment for on page 34 he says "I should put the conditional probability that if a simple law exists it is of the form $a+b\cdot c^n$ somewhere between 1/8 and 1/3". The important word here is "simple", which implies a familiar mathematical sequence occurring frequently elsewhere in nature.

If we accept Good's subjective assessment of the simple laws of cosmology as consisting of proportion 1/8 to 1/3 of the family $a+b\cdot c^n$, then we should reduce his final likelihood ratio by a factor of 1/8 to 1/3, since the remaining 7/8 to 2/3 of the simple laws cannot contribute much likelihood to the observed data (e.g. a nearly geometric sequence isn't also nearly quadratic). Good does not make such a reduction nor does he even allow for the special choice of $c=2$ in the form $a+b\cdot c^n$.

As mentioned before, Good does assess a penalty factor of 5 against Bode's law for failing in the cases $n=1$, 9, and 10. One might adopt the harsh viewpoint that there were potentially $\binom{10}{3}=120$ Bode's laws of the form $a+b\cdot c^n$, one for each choice of three exceptions to the rule.

(What if the three nearest or three farthest or three heaviest planets had not fitted the formula? Wouldn't we still be considering some form of Bode's law?) Although I personally feel that 5 is too small a penalty factor for this situation (and 120 too large) I will use Good's figure for comparison purposes at the end of the analysis in Section 3.

Criticism 3. Nuisance Parameters.

The prior distributions Good assigns to $a, b, \sigma,$ and $\log D/d$ are designed to express ignorance rather than any actual knowledge of the generating mechanism for these parameters. (Good is not claiming to know, for instance, that 50% of those planetary systems having second through eighth planets not obeying Bode's law are too small to even include the observed orbits of Venus through Uranus.)

The usual assurances that the particular choice of a vague prior distribution isn't important to the final conclusion are suspect here.* The trouble is that model B and model $\bar{B}$ have completely different structures, with dissimilar likelihood functions expressed in terms of different sets of parameters. If we let the prior distribution on the parameter a get more and more diffuse without changing the other priors the likelihood ratio of B to $\bar{B}$ is reduced to zero; or we can make this same ratio arbitrarily large by making the prior on $\log D/d$ suitably diffuse. Good does not present any evidence that it

---

*Good does not make any such assurances. Quoting from another paper of Good's [4], page 402, "... estimates are usually less sensitive than are significance tests to the precise initial distribution assumed".

is "fair" to assume that all four nuisance parameters have the same relative prior distributions around their maximum likelihood values. (From the fact that the Neyman-Pearson likelihood ratio is about 3 times larger than Good's Bayesian likelihood factor a non-Bayesian might feel that these priors are mildly biased against Bode's law.)

In any case it would be preferable not to have to make assumptions about nuisance parameters. The analysis of Section 3 sidesteps this pitfall by dealing with the relative spacings of the planets, rather than their distances in astronomical units.

3. Another Model for the Planetary Distances.

We will assume the following model called "C" (for "convex") for the situation where Bode's law does not hold:

$$C: \text{ Let } U_i = \frac{D_{i+2}-D_{i+1}}{D_8-D_2} \quad \text{for } i = 1, 2, \ldots, 6 .$$

Then under the hypothesis C that Bode's law does not hold the random variables $U_1, U_2, \ldots, U_5$ have joint density

$$f(u_1, u_2, \ldots, u_5) = 6!5! \quad \text{for } 0 \leq u_1 \leq u_2 \ldots \leq u_5 \leq 1 - \sum_1^5 u_i .$$

(That is $U_1, U_2, \ldots, U_5$ and $U_6 = 1 - \sum_{i=1}^{5} U_i$ have the joint distribution of a random division of the interval $[0, 1]$ into 6 segments, subject to the condition that the segments are in increasing order of length.)

Here C will play the role of the "alternative less interesting statistical model" labelled (ii) in Section 1.* Its choice is an obvious attempt to avoid criticism 1 of Section 2. Since we are going to use a standard "objective" significance test of C, in Fisher's sense of the term, we do not have to supply a specific model (i) but only an objective statistic that measures in some way the agreement of a random vector $(U_1, U_2, \ldots, U_6)$ with the hypothesis "the distances follow some simple law".

Let us forget criticism 2 for a moment and assume that the only simple laws are geometric ones. If the distances were exactly geometric, that is if for some $c > 1$ the distances were of the form $d_n = a+b \cdot c^n$ for $n=2,3,\ldots,8$, then the observed spacings

$$u_i = (d_{i+2} - d_{i+1})/(d_8 - d_2)$$

would satisfy

$$u_i = cu_{i-1} \quad \text{for} \quad i = 2, 3, \ldots, 6 ,$$

or equivalently

$$u_i = c^{i-1}/s_c \qquad i = 1, 2, \ldots, 6 ,$$

*Ordinarily C would be called the null hypothesis, but we have avoided this term so far in deference to [3] (c.f. the last sentence of Anscombe's remarks on Good's paper [3]). Sections 4 and 5 and the appendix use this more standard nomenclature.

where

$$s_c = \frac{c^6 - 1}{c - 1} .$$

The objective statistic should be chosen to measure agreement with -- or, equivalently, distance from -- this ideal. We will use the distance statistic

$$\Delta(u_1, u_2, \ldots, u_6) = \inf_{1<c<\infty} \left[\sum_{i=1}^{6} \frac{(u_i - c^{i-1}/s_c)^2}{c^{i-1}/s_c} \right]^{\frac{1}{2}} . \qquad (1)$$

(Some justification for this particular choice is given in Section 5.)

The spacings $u_1, u_2, \ldots, u_6$ for the observed planetary distances as given in table 1 are computed to be .01499, .02836, .07447, .1246, .2350, and .5225 respectively. Numerical minimization yields

$$\Delta(u_1, u_2, \ldots, u_6) = .0646$$

for the data, the minimizing value of c being 2.016.

The hypothesis C was simulated on a computer, and 2000 independent realizations $(u_1, u_2, \ldots, u_6)$ of the vector $(U_1, U_2, \ldots, U_6)$ were generated. For each one $\Delta(u_1, u_2, \ldots, u_6)$ was calculated, along with the minimizing value of c. The results are partially summarized in figure 1. It can be seen that the observed value $\Delta = .0646$ has a tail probability (significance level) of about 1%. There were 22 out of the 2000 simulated cases with $\Delta < .0646$. (The smallest recorded value was $\Delta = .0138$.) Figure 2 graphically displays the minimizing values of c corresponding to the small values of Δ.

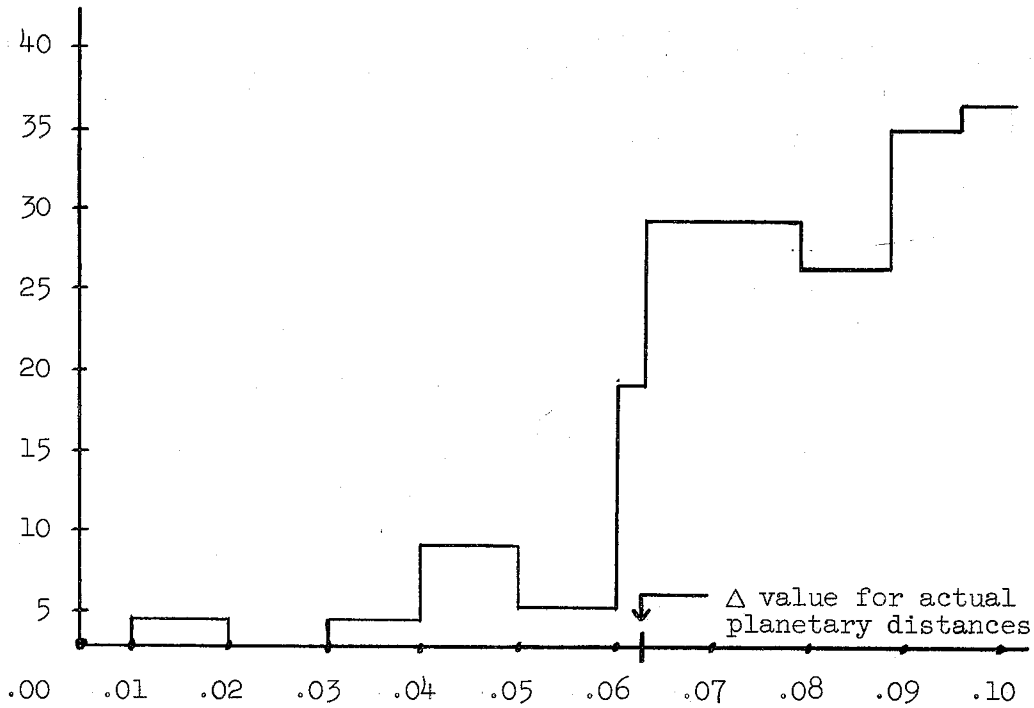

Figure 1. Histogram of 2000 △ values.

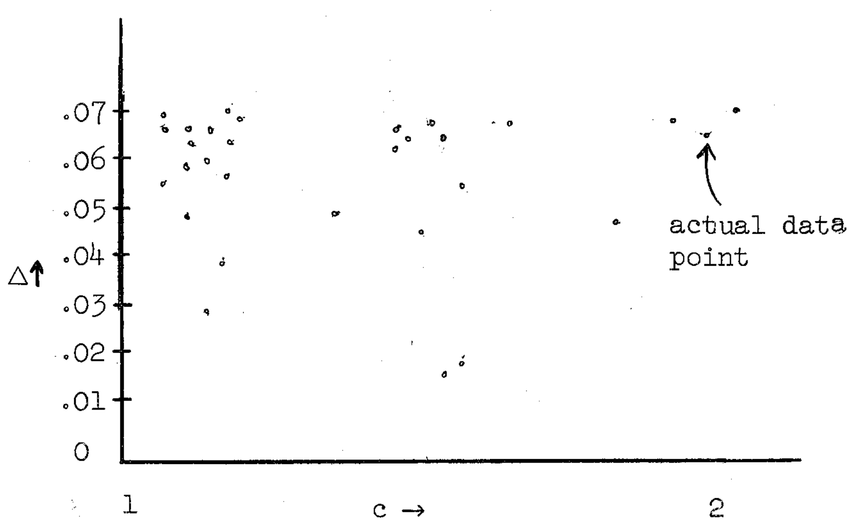

Figure 2. Minimizing values of c plotted against corresponding values of △ (△ < .07).

If we wish to answer criticism 2 of Good's model we must consider forms of "simplicity" other than the geometric progression. The quadratic model mentioned in Section 2 is a natural candidate,* and yields an arithmetic progression for the spacings. The family of all such arithmetic progressions is given as a function of $c > 0$ by

$$u_i = \frac{1+(i-1)c}{s_c^*}, \quad i = 1, 2, \ldots, 6,$$

where

$$s_c^* = 6 + 15c .$$

(This is the family of spacings that results from the family of distances

$$d_n = A + Bn + Cn^2 ,$$

subject to the constraints that both d_n and $d_{n+1}-d_n$ are increasing functions of n.)

If we use the same concept of distance as before, the distance from an observed vector $(u_1, u_2, \ldots, u_6)$ to the additive family is

$$\Delta^*(u_1, u_2, \ldots, u_6) = \inf_{0<c} \left[\sum_{i=1}^{6} \frac{\left(u_i - \frac{1+(i-1)c}{s_c^*}\right)^2}{\frac{1+(i-1)c}{s_c^*}} \right]^{\frac{1}{2}} . \qquad (2)$$

*It will be obvious from what follows that this particular choice is not essential to our conclusion.

An additional 1000 random choices of $(u_1, u_2, \ldots, u_6)$ according to the model C yielded 12 cases with $\Delta^* < .0646$, again about 1% of the sample.

Finally another 1000 random vectors $(u_1, u_2, \ldots, u_6)$ were generated according to C, and the statistic

$$\Delta^{**} = \min(\Delta, \Delta^*) \tag{3}$$

was calculated for each. This series yielded 24 cases with $\Delta^{**} < .0646$, the observed value for the planetary spacings. (Of these 12 had $\Delta < \Delta^*$ and 12 had $\Delta^* < \Delta$.) This shows that the events "Spacings closer than .0646 to geometric" and "Spacings closer than .0646 to arithmetic" are effectively disjoint under C, each with probability about 1%.

We can paraphrase the results so far as follows: "the model C yields spacings as simple (in either a geometric or arithmetic sense) as the actual planetary spacings in about 2% of the cases. Now we could continue adding laws to our definition of simplicity until the percentage of cases under C where the minimum distance was less than .0646 was arbitrarily large. What stops us is a subjective feeling that these laws really wouldn't be as simple as the geometric or arithmetic progressions, and shouldn't receive equal weighting with these in assessing how close a randomly generated vector $(u_1, u_2, \ldots, u_6)$ is to "simplicity". Let us accept Good's most optimistic assessment that the geometric progressions are 1/3 of all the simple laws. From the results above we then feel that we should increase our assessment of the probability of the event "A vector randomly generated by C is simpler than the actual

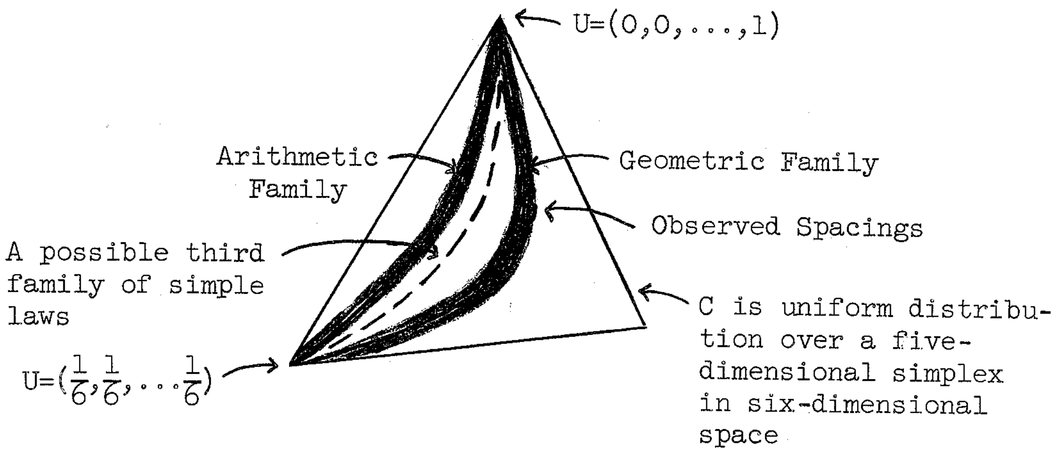

Figure 3. Schematic Representation of C and simple laws. Shaded regions indicate points closer than .0646 to geometric or arithmetic families.

planetary spacings" from .02 to about .03. (This assumes that any other one-parameter family of simple laws we would consider would have about the same tail probability, .01, for the event of interest, disjointly from the geometric and arithmetic families. Crude geometric considerations, as illustrated in figure 3, lend credance to the validity of this assumption, at least at the level of approximation we are tolerating here.)

To compare our results with Good's we must first transform the tail probability .03 into a likelihood ratio (in his terminology a "Bayes Factor") for B to C. In other papers [4, 5] Good has suggested that a reasonable transformation is

$$\text{Likelihood Ratio} = \gamma \frac{1}{\text{Tail Probability}},$$

where the factor γ is between $\frac{3}{10}$ and $\frac{1}{30}$. We will use the factor $\gamma = \frac{1}{4}$ for reasons stated in the appendix. This gives the likelihood ratio $\frac{1}{4}\,\frac{1}{.03} \approx 8$ in favor of Bode's law over C. When we finally adjust by Good's penalty factor of $\frac{1}{5}$, to compensate for ignoring Mercury, Neptune, and Pluto, the likelihood ratio is reduced to nearly unity.

The conclusion is that if we believe the weaker "law of increasing differences", then there is no compelling evidence for believing that Bode's law is not artifactual.

The next two sections are devoted to discussions of the two crucial points in an analysis: the choice of the null hypothesis, and the choice of the objective statistic. In the appendix we discuss Good's interesting conversion factor from tail probability to likelihood ratio.

4. Choosing a Null Hypothesis (And a Defense of Numerology).

In the spirited discussion which followed the presentation of his paper [3] Professor Good was accused of practicing "numerology" in his Bode's law investigation.* In principle, testing the validity of Bode's law is not an illegal or even a particularly unusual use of statistical theory. We are asked to consider a hypothesis postulated after the data was partially collected, but experience shows this to be the norm rather than the exception. We have difficulty defining exactly what *is* the hypothesis of interest (i.e. what is a simple rule), but we face this difficulty every time we choose between a one-tailed and a two-tailed test.

*This is a serious charge against a statistician, approximately equivalent to detecting religious tendencies in a theologian.

The most interesting feature of this problem is the light it casts on the role of the null hypothesis in hypothesis testing. These terms are used here in Fisher's sense in which the null hypothesis is by design a hypothesis of uninteresting structure compared to that which we are considering as an alternative, though it may contain its own interesting features. For example, the "law of increasing differences" might be considered interesting in its own right. If so, and if its validity didn't seem apparent to the scientists involved, then it itself could be compared with a still less interesting, more null hypothesis, e.g. Good's $\bar{B}$. As was indicated in Section 2 such a test would probably give reasonable grounds for believing in increasing differences.

However it is not necessary to believe in the null hypothesis in order to use it as a test against the alternative of interest. Very often, perhaps most of the time, we do <u>not</u> believe in the validity of the Fisherian null hypothesis, whether or not the test based on it accepts or rejects in the usual sense. For example, having observed 15 drier-than-average and 10 wetter-than-average summers in Seattle since 1945, one needn't believe in a coin tossing model in order to give some statistical reassurance that nuclear testing has not modified Seattle's weather.

The null hypothesis in the context we are discussing plays the role of devil's advocate, a competitor that an alternative of interest to us must soundly discredit in order to show its strength.* The standard statistical situations (one-sample problems, two-sample problems, etc.) have given rise to standard null hypotheses of more or less agreed upon potency. Perhaps the disturbing aspect of the Bode's law problem is that Professor Good has pointed out an interesting hypothesis testing situation that does not have a standard repetoire of null hypotheses. One purpose of this paper is to provide a small beginning for such a repetoire.

The conclusions of a significance test are bound to be less than completely satisfying given the indirect route of the argument. In the case at hand for instance, accepting C doesn't mean we believe C is true (figure 2 mildly discourages such a belief). All we can say is that a statistical model that is relatively uninteresting compared to Bode's law would often yield data as "simple" as that actually observed, and this undermines the necessity of our belief in the law's validity. Conversely even if we had decisively rejected C we still might fear that we had overlooked some other reasonable null hypothesis which would do better.

One should not be dismayed by the limitations of the Fisherian significance test since it is designed only to give us some direction toward the correct answer in situations like the present one where there

*Anscombe [1] puts it this way: "A secondary use (of significance tests), destructive in intention, is to demonstrate that some rudimentary hypothesis is inadequate to explain the observations, so that there is a prima facie case for opening an investigation ... No doubt this was the use Fisher particularly had in mind in suggesting the term 'null hypothesis'." Anscombe's "primary use" of significance tests is in the Pearsonian goodness-of-fit manner.

is little data to work with. As more data accumulates in any given problem significance testing becomes superflous. We would certainly not be using the present analysis if we had measurements on 50 solar systems. If we were still at all interested in Bode's law by then it would probably be because it had an obvious gross validity, and we would be estimating the deviations of the measurements from the ideal. By definition "estimation" refers to situations where we believe we know all the possible relevant statistical models and we are simply trying to choose the correct one.* Estimation is an inherantly more satisfying operation than significance testing, but demands more data or more theoretical knowledge from the statistician.

5. Choice of the Distance Function.

The objective statistic, or distance function Δ, in our analysis is supposed to measure the distance of an observed sequence $d_2, d_3, \ldots, d_8$ from "simplicity". The smaller the observed distance the greater the simplicity, and the stronger the evidence against our null hypothesis. If we believe that any such distance function must be invariant under linear transformations

$$\tilde{d}_i = A + Bd_i \qquad B \neq 0,$$

*Following Fisher's terminology, this includes hypothesis testing models such as the analysis of variance where we are trying to decide whether or not some parameters are zero. Anscombe points out in [1] that the term hypothesis test has been used to cover an unfortunately large number of distinct situations.

then it is possible to express Δ in terms of the spacings as we did in Section 3.

The statistic Δ^{**} used in Section 3 can be thought of as a special case, with $p=1$, of the family of distance functions

$$\Delta_p(u_1, u_2, \ldots, u_6) = \inf_{\xi\in\Xi} \left[\sum_{i=1}^{6} \frac{(u_i - \tau_i(\xi))^2}{(\tau_i(\xi))^p} \right]^{\frac{1}{2}} . \qquad (4)$$

Here ξ indexes a family of partitions Ξ of the unit interval into ordered segments

$$0 < \tau_1(\xi) \leq \tau_2(\xi) \leq \cdots \leq \tau_6(\xi) < 1 , \quad \sum_{i=1}^{6} \tau_i(\xi) = 1 .$$

(For example at (1) of Section 3 we could take $\tau_i(\xi)=\xi^{i-1}/[(\xi^6-1)/(\xi-1)]$ for $\xi\in(1,\infty)$, while at (3) ξ would have to index both the geometric and additive family of partitions.) The choice $p=0$ gives equal weight to errors $|u_i-\tau_i(\xi)|$ regardless of the size of $\tau_i(\xi)$, while the choice $p=2$ essentially measures the <u>relative</u> error $|u_i-\tau_i(\xi)|/\tau_i(\xi)$ of each segment.

The value $p=1$ was selected for the actual analysis on the purely subjective grounds that it was a compromise between $p=2$, which seemed to overrate the importance of the smallest intervals, and $p=0$ which erred in the other direction. (As a hedge 600 more random vectors obeying model C were generated, and their $p=2$ distances $\Delta_2(u_1,u_2,\ldots,u_6)$ from the geometric family tabulated. This yielded 10 distances less than .226, the observed Δ_2 value for the actual planetary spacings, with an estimated tail probability of 10/600, slightly larger than the .01 value for $p=1$. A similar experiment with $p=0$ yielded 5 out of 500 cases with $\Delta_o < .023$, the Δ_o value for the actual distances.)

The choice of a distance function for our hypothesis test can be made less arbitrary if we are willing to provide a statistical model for what we mean by Bode's law being true. (Following Good we will label any such model "B", with subscripts to indicate different formulations.) For example let $B_{\Sigma(\xi)}$ be the hypothesis that the random vector $U=(U_1,U_2,\ldots,U_6)$ is normally distributed with mean vector $\tau(\xi)=(\tau_1(\xi),\tau_2(\xi),\ldots,\tau_6(\xi))$ and covariance matrix $\Sigma(\xi)$ for ξ in some index set Ξ,

$$B_{\Sigma(\xi)}: \quad U \sim \mathcal{N}(\tau(\xi), \Sigma(\xi)), \quad \xi \in \Xi .$$

Then the Neyman-Pearson likelihood ratio of C to $B_{\Sigma(\xi)}$ is a decreasing function of

$$\Delta_{\Sigma(\xi)}(u_1, u_2, \ldots, u_6) = \inf_{\xi\in\Xi}\{[(u-\tau(\xi))\Sigma^{-1}(\xi)(u-\tau(\xi))']^{\frac{1}{2}}+\log|\Sigma(\xi)|\}. \quad (5)$$

(Since the vector U is subject to the linear constraint $\sum_1^6 U_i=1$ the "inverse" $\Sigma^{-1}(\xi)$ in this definition is actually any pseudo-inverse of the singular matrix $\Sigma(\xi)$, [6], and $|\Sigma(\xi)|$ is the product of the five non-zero eigenvalues of $\Sigma(\xi)$.)

The statistic Δ_1 is obtained if $\Sigma^{-1}(\xi)$ is diagonal with entries along the main diagonal proportional to $1/\tau_i(\xi)$, and we ignore the $\log|\Sigma(\xi)|$ term in (5). (Ignoring this term is formally similar to working with X^2 rather than maximum likelihood in a multinomial goodness-of-fit problem, an approximation which ordinarily has little effect on the defined statistic.) The diagonal matrix with entries $1/\tau_i(\xi)$ is the pseudo-inverse of the matrix

$$\Sigma(\xi) = \begin{pmatrix} \tau_1(\xi)-\tau_1^2(\xi) & -\tau_1(\xi)\tau_2(\xi) & \cdots & -\tau_1(\xi)\tau_6(\xi) \\ -\tau_1(\xi)\tau_2(\xi) & \tau_2(\xi)-\tau_2^2(\xi) & \cdots & \vdots \\ \vdots & & & \vdots \\ -\tau_1(\xi)\tau_6(\xi) & \cdots & & \tau_6(\xi)-\tau_6^2(\xi) \end{pmatrix}, \quad (6)$$

a fact which is useful in the theory of the multinomial distribution [6]. We could expect approximately this covariance matrix for U if the planetary separations $D_{n+1}-D_n$ were independent random variables with variances proportional to their means. (This statement follows from standard "delta method" arguments, and for these to be convincing the proportionality constant mentioned must be small enough so that each separation has a small coefficient of variation.)

Physically we do not expect the planetary separations to be independent of one another, and it might be argued that this is even less likely than Good's assumption that the distances themselves are independent. Of course we are free to use any statistic we want, in particular Δ_1, even if we are not certain of the rectitude of the model B underlying its choice.* As payment for this freedom we expect to suffer a loss of

*We may not even believe that Bode's law, if true, can be stated in terms of a statistical model. Using a plausible model B to suggest a hypothesis testing statistic is similar in spirit to the "Bayes/non-Bayes compromise" suggested by Good in Section 3.4 of [3]. Here it would be a "Fisher/Neyman-Pearson compromise".

power if Bode's law is true in some form $B_{\Sigma(\xi)}$, but the statistic Δ_1 we are using is far different from the appropriate statistic $\Delta_{\Sigma(\xi)}$. The discussion which follows is designed to give a rough idea of the power loss.

First suppose that we wish to test the null hypothesis H_o that an N-dimensional random vector X is uniformly distributed over some N-dimensional region $\mathfrak{X}$ in Euclidean N-space,

$$H_o: \quad f^X(x) = \frac{1}{\mathrm{Vol}(\mathfrak{X})} \qquad x \in \mathfrak{X}$$
$$= 0 \qquad x \notin \mathfrak{X} \tag{7}$$

versus the alternative that X is normally distributed with mean vector $\mu \in \mathfrak{X}$,

$$H_1: \quad X \sim \mathfrak{N}(\mu, \Sigma) ,$$

where Σ is an N by N matrix of full rank, unknown to us.

Let Λ be any N by N matrix of full rank, and define the <u>expected significance level</u> of the hypothesis test of H_o versus H_1 based on the statistic $\Delta_\Lambda=[(X-\mu)\Lambda^{-1}(X-\mu)']^{\frac{1}{2}}$ as the expected value under H_1 of the H_o tail probability of Δ_Λ. We can write this as

$$\mathrm{ESL}(\Delta_\Lambda) = \mathrm{Prob}\{\Delta_\Lambda^{(o)} \leq \Delta_\Lambda^{(1)}\} ,$$

where $\Delta_\Lambda^{(o)}$ and $\Delta_\Lambda^{(1)}$ are independent Δ_Λ values generated under H_o and H_1 respectively. By definition Δ_Λ is a powerful statistic for testing H_1 against H_o if $\mathrm{ESL}(\Delta_\Lambda)$ is small. The Neyman-Pearson lemma can be used to show that the likelihood ratio statistic Δ_Σ

minimizes the ESL among <u>all</u> statistics, including of course the class of quadratic forms we are considering here [2]. (To be more precise, $\Delta_{\not\Sigma}$ is the likelihood ratio statistic if X is observed in $\mathfrak{X}$. Let us add to our assumptions the condition that $\mathfrak{X}$ be large compared to the variation of X under H_1, so we do not have to worry about "fringe effects" in computing the ESL. It is easy to make this condition rigorous, but the expense in notation belies the crude application of the lemma below.)

If we knew $\not\Sigma$ we would certainly base our test on $\Delta_{\not\Sigma}$, but if not we can at least compute how much the wrong choice of statistic costs us in ESL:

Lemma. Under the conditions stated

$$\frac{ESL(\Delta_\Lambda)}{ESL(\Delta_{\not\Sigma})} = \frac{\Gamma(\frac{N}{2})}{2^{N/2}\Gamma(N)} \; \frac{E\left\{\sum_{i=1}^{N} \beta_i S_i\right\}^{N/2}}{\sqrt{\prod_{i=1}^{N} \beta_i}} ,$$

where the S_i are independent χ_1^2 random variables, and the β_i are the eigenvalues of $\not\Sigma\Lambda^{-1}$.

Proof: By first applying standard linear transformations we can assume without loss of generality that $\mu=0$, $\not\Sigma=I$, and Λ is a diagonal matrix with diagonal entries $1/\beta_i$ (see [6], page 37). Then

$$\text{Prob}\left\{\Delta_\Lambda^{(0)} \leq \Delta_\Lambda^{(1)} \middle| \Delta_\Lambda^{(1)} = \Delta\right\} = \frac{k_N}{Vol(\mathfrak{X})} \; \frac{\Delta^N}{\sqrt{\prod_{i=1}^{N} \beta_i}} ,$$

where $k_N = 2\pi^{N/2}/(N\Gamma(N/2))$ is the proportionality constant relating the volume of an N-dimensional sphere to the N^{th} power of its radius. (This assumes that the ellipsoid $\{x: \sum_{i=1}^{N} \beta_i x_i^2 \leq \Delta^2\}$ is entirely contained in $\mathscr{X}$.) Since under H_1 the quantity $\Delta^N = (\sum_{i=1}^{N} \beta_i X_i^2)^{N/2}$ is distributed as $(\sum_{i=1}^{N} \beta_i S_i)^{N/2}$ with the S_i independent χ_1^2 random variables we have

$$\mathrm{ESL}(D_\Lambda) = \frac{k_N}{\mathrm{Vol}(\mathscr{X})} \; \frac{E\left(\sum_{i=1}^{N} \beta_i S_i\right)^{N/2}}{\sqrt{\prod_{i=1}^{N} \beta_i}} \; .$$

Likewise

$$\mathrm{ESL}(D_I) = \frac{k_N}{\mathrm{Vol}(\mathscr{X})} E\left(\sum_{i=1}^{N} S_i\right)^{N/2} = \frac{k_N}{\mathrm{Vol}(\mathscr{X})} \; \frac{2^{N/2}\Gamma(N)}{\Gamma(N/2)} \; ,$$

completing the proof.

To return to the analysis of the planetary distances, notice that our null hypothesis C is a uniform distribution over a subset of a 5 dimensional linear space, and that if Bode's law is represented in some form $B_{\not{\Sigma}(\xi)}$, then the alternative hypothesis is described by various one-dimensional curves of "simple" mean vectors $\tau(\xi)$ through this space. We can apply the lemma above if we are willing to accept the familiar reduction from a one-dimensional hypothesis in 5-space to a zero-dimensional hypothesis in 4-space. Applying the lemma with $N=4$ gives

$$\frac{\mathrm{ESL}(\Lambda)}{\mathrm{ESL}(\not{\Sigma})} = \frac{\left(\sum_{i=1}^{4} \beta_i\right)^2 + 2\sum_{i=1}^{4} \beta_i^2}{24\sqrt{\beta_1\beta_2\beta_3\beta_4}} \; . \qquad (8)$$

(Here the β_i are the four non-zero roots of $\Sigma(\xi)\Lambda^{-1}(\xi)$ after we have adjusted for the reduction in dimension from five to four. The β_i are functions of ξ, but for (8) to be useful as an approximation the estimated values $\beta_i(\hat{\xi})$, where $\hat{\xi}$ is the "Λ" minimum distance estimate of ξ, must have small coefficients of variation under $B_{\Sigma(\xi)}$. There is no approximation involved in going from the lemma to (8) if $\Sigma(\xi)$ and $\Lambda(\xi)$ are not functions of ξ, and $\tau(\xi)$ is a linear subspace of C.)

The interesting feature of equation (8) is that it is not very sensitive to deviations of the β_i values from the ideal $\beta_i=1$, $i=1,2,3,4$. For example suppose we observed a four-dimensional vector X and we were using ordinary Euclidean distance $\Delta_I = \sum_{i=1}^{4}(X_i-\mu_i)^2$ to test the null hypothesis that X was uniform over some large region $\mathcal{X}$ containing the vector μ, versus the alternative that X had mean μ and small variation about μ. Furthermore suppose that the mean of X was indeed μ, but that the X_i were actually successive differences of five independent normal random variables, and so were normal with covariance matrix

$$\Sigma = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix} .$$

It can be shown [7] that in this case

$$\beta_i = 2 - 2 \cos \frac{i\pi}{5} \qquad i = 1, 2, 3, 4,$$

which yields from equation (8)

$$\frac{ESL(I)}{ESL(\Sigma)} = 2.01 \ .$$

That is, it costs us a factor of 2 in ESL to behave as if the components X_i were independent when in fact they are successive differences of independent variables.

In Good's model B for Bode's law the planetary distances are assumed to be independent, while our analysis tacitly assumed that the differences of the distances were independent. The discussion above was intended to show, by admittedly imperfect analogy, that this is probably not the source of disagreement between our conclusion and his.

References

1. Anscombe, F. J. (1962) "Tests of Goodness of Fit", JRSS Series B, Vol. 25, No. 3, 81-94.

2. Dempster, A. P. and Schatzoff, M. (1965) "Expected significance level as a sensitivity index for test statistics", JASA, Vol. 60, 420-436.

3. Good, I. J. (1969) "A subjective evaluation of Bode's Law and an 'Objective' Test for approximate numerical rationality", with discussion by H. O. Hartley, I. J. Bross, H. A. David, M. Zelen, R. E. Bargmann, F. J. Anscombe, M. Davis and R. L. Anderson. JASA, Vol. 64, March, 23-66.

4. Good, I. J. (1967) "A Bayesian significance test for multinomial distributions", (with discussion) JRSS Series B, Vol. 29, No. 3, 399-431.

5. Good, I. J. (1965) The Estimation of Probabilities, Cambridge, Mass., M.I.T Press, p. 35.

6. Rao, C. R. (1965) Linear Statistical Inference and Its Applications New York, John Wiley and Sons.

7. Von Neuman, J. (1941) "Distribution of the ratio of the mean square successive difference to the variance", Annals of Math. Stat., Vol. 12, 367-395.

Appendix: The Relationship of Likelihood Ratios to Tail Probabilities.

Good [2, 3] has suggested that in many situations the observed tail probability

$$P(\triangle) = \text{Prob}_{H_o}\{\Delta \leq \triangle\}$$

of a statistic Δ which is being used* to test a null hypothesis H_o is related to the observed likelihood ratio of H_o versus an alternative hypothesis H_1,

$$\ell(\triangle) = f_{H_1}^{\Delta}(\triangle)/f_{H_o}^{\Delta}(\triangle)$$

by the relationship

$$\ell(\triangle) = \gamma(\triangle)\,\frac{1}{P(\triangle)}\,, \qquad (1A)$$

where the factor γ satisfies

$$\frac{1}{30} \leq \gamma \leq \frac{3}{10}\,.$$

This "γ factor" is intended to apply to cases where both H_o and H_1 are simple, or are reduced to simplicity by Bayesian methods as in Good's analysis of Bode's law, and where $P(\triangle)$ is in the usual significance testing range, say $P(\triangle) \leq .05$.

---

*We are continuing the convention of the paper whereby all test statistics are non-negative, with small observed values being considered evidence against H_o.

If we recall that the "risk figure" for the family of Neyman-Pearson likelihood ratio tests of H_o versus H_1 has slope $-\ell$ at the point (α,β) corresponding to the test which rejects for observed likelihood ratios $\geq \ell$, we see that if Δ is any monotone (decreasing) function of the likelihood ratio then γ has the geometrical interpretation shown in figure 1A.

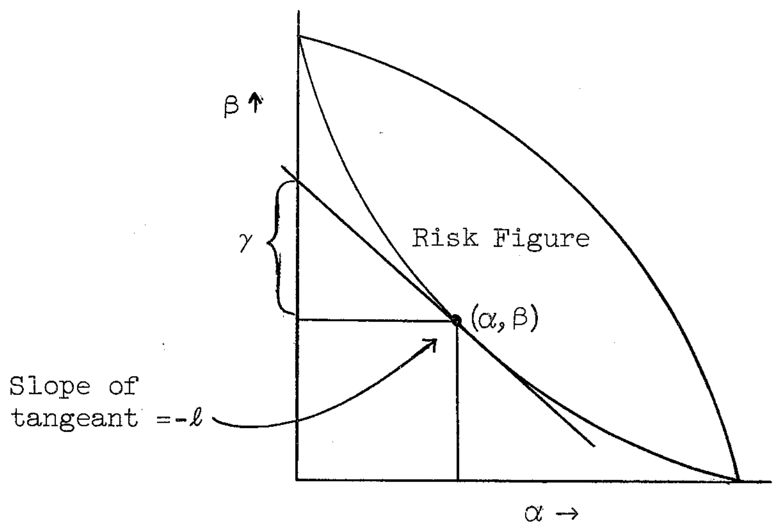

Figure 1A. Relationship of Good's γ Factor to Neyman-Pearson Risk Figure.

This shows that

(1) $\gamma \leq 1$ (more precisely $\gamma = \alpha\ell \leq \min(1-\beta, \ell)$).

(2) γ can be made arbitrarily small for any fixed value of α by considering situations with arbitrarily high discrimination between H_o and H_1 (i.e. experiments where we can achieve the desired α-level with an arbitrarily low value of ℓ).

(3) Good's bound on the factor γ is a statement about the shape of risk figures for statistically reasonable values of α and β.

We will now calculate some properties of γ for the situation described in section 5, where the observed vector X has either the null distribution (6) or the alternative distribution (7). (However, we shall assume that we know Σ as well as μ so that by linear transformations we can take $\mu=0$ and $\Sigma=I$.)

Our statistic will be

$$\Delta(X) = \frac{\|X\|^2}{2},$$

so that rejecting for small observed values of Δ, $\Delta \leq \Delta_o$, is a likelihood ratio test of H_o versus H_1. Under these conditions we have the following

<u>Theorem</u>: $\gamma(\Delta) = \Delta^{N/2} e^{-\Delta}/\Gamma(N/2+1)$ for all values of Δ such that the sphere $\{X: \|X\|^2 \leq 2\Delta\}$ is entirely contained in $\mathfrak{X}$. The maximum value of γ is

$$\hat{\gamma} = \left(\frac{N}{2}\right)^{N/2} e^{-N/2}\Big/\Gamma(N/2+1)$$

attained at $\hat{\Delta}=N/2$, while the expected value of $\gamma(\Delta)$ under H_1 is

$$E_{H_1}\gamma(\Delta) = \frac{1}{2^N}\frac{\Gamma(N)}{\Gamma(N/2+1)\Gamma(N/2)}.$$

(The maximum and expectation are calculated ignoring the fact the expression given for $\gamma(\Delta)$ does not hold for large Δ.)

Proof: If the sphere $\{x: \|X\|^2 \le \Delta\}$ is contained in $\mathscr{X}$ then

$$P(\Delta) = \text{Prob}_{H_0}(\boldsymbol{\Delta} \le \Delta) = \frac{k_N (2\Delta)^{N/2}}{\text{Vol}(\mathscr{X})} ,$$

where $k_N = (2\pi^{N/2})/(N\Gamma(N/2))$ as before. We also calculate

$$\ell(\Delta) = (2\pi)^{-N/2} e^{-\Delta} \cdot \text{Vol}(\mathscr{X}) ,$$

and so

$$\gamma(\Delta) = P(\Delta)\ell(\Delta)$$

$$= \frac{\Delta^{N/2} e^{-\Delta}}{\Gamma(N/2+1)} .$$

This attains its maximum at $\hat{\Delta}=N/2$. Under H_1 the distribution of $\boldsymbol{\Delta}$ is $\frac{1}{2}\chi^2_N$, and the formula for the expectation of $\gamma(\Delta)$ is obtained as the expectation of $\Delta^{N/2}e^{-\Delta}/\Gamma(N/2+1)$ under this distribution.

For the case $N=4$ appropriate to the Bode's law problem we calculate

$$\hat{\gamma} = .27, \quad E_{H_1}\gamma(\boldsymbol{\Delta}) = .1875 .$$

The value $\gamma=\frac{1}{4}$ used in section 3 is seen to be close to the maximum value of γ possible for the situation above.

If we only know Σ up to a proportionality constant, then the alternative hypothesis can be reduced to $X \sim \mathcal{N}(0, \sigma^2 I)$, with σ^2 unknown.

The theorem still holds as stated, except that the formula for $\gamma(\Delta)$ becomes

$$\gamma(\Delta) = \frac{(\Delta/\sigma^2)^{N/2} e^{-\Delta/\sigma^2}}{\Gamma(N/2+1)} ,$$

and the maximizing value is $\hat{\Delta} = \frac{N}{2}\sigma^2$. (Notice that if we estimate σ^2 with its (H_1) maximum likelihood estimate $\hat{\sigma}^2 = \frac{\|X\|^2}{N}$, then $\hat{\Delta} = \frac{N}{2}\hat{\sigma}^2 = \frac{\|X\|^2}{2} = \Delta$, the observed value. In other words the maximum likelihood estimate of $\gamma(\Delta)$ is its maximum possible value.) In particular the maximum and expected value (under H_1) of γ are as stated in the theorem, and do not depend on σ^2. This implies that if we have any prior distribution on σ^2 the marginal expected value of γ, which is the same as the expected value of γ when H_1 is reduced to a simple hypothesis by integrating out σ^2, will be as given (up to the approximation noted after the theorem).