ON THE PRESENCE OF TRENDS IN "RANDOM" SEQUENCES

by

Donald A. Berry

Technical Report No. 239

University of Minnesota

Minneapolis, Minnesota

1. Introduction.

An experiment is conducted to arrange the numbers 1,2,3,4,5 and the result is the sequence (5,4,3,2,1). One frequently hears the question : "Is (5,4,3,2,1) a random arrangement?" Such a question is meaningless. The questioner may have meant: "In producing the arrangement (5,4,3,2,1) was the mechanism random?" While more meaningful, this question still cannot be answered. One needs to know something about the mechanism, and no particular arrangement provides such information.

Consider the controversy surrounding the 1970 U. S. draft lottery; see the thorough discussion and careful analysis in (Fienberg 1971). Some people argued that since having a birthday late in the year meant one was more likely to be high on the drafting list, the lottery could not have been random. Such post hoc reasoning can easily be fallacious. For every sequence it is at least a little better to be born either early or late! The point is that every arrangement seems "nonrandom" in some respect; some less random than others. If the mechanism is in fact random (or, more realistically, can reasonably be assumed random) then every arrangement is as likely as any other and so, in a sense, every arrangement is "random"; certainly no arrangement is less "random" than another. This statement contradicts the feeling one has that, for example, (4,1,3,5,2) is "random" while (1,2,3,4,5) is not. The basis for such feeling probably lies in the fact that one frequently encounters the latter arrangement as the natural ordering of the first five positive integers in contexts separate from questions of randomness. If a person associates some probability with the possibility that the five numbers were arranged in natural order (and therefore the mechanism is not random) then that probability would be substantially increased if the arrangement (1,2,3,4,5) were observed and, of course, would be annihilated if any other arrangement were observed. For such a person the arrangement (1,2,3,4,5) may deserve to be called "less random" than (4,1,3,5,2), for example.

In any problem "randomness" is only one of many models that could be considered. It usually plays a special role in statistics in part because the calculations are relatively easy and the mathematics well understood. One usually makes inferences assuming randomness and then worries about whether the assumption is true, and what effect different kinds of non-randomness will have on the inferences. For some, a natural question to ask is: "What is the probability that the randomness assymption is correct?" Such a question can only be answered using Bayes' theorem, and only when particular alternatives have been proposed. It is always easy to propose alternatives which make the randomness assumption look bad; one such is the model that predicts that the only observation (or set of observations) possible is the one actually obtained! Speaking somewhat loosely, it seems fair to consider a family of alternatives each member of which is no more specific than the randomness assumption. The remainder of this paper will address the question posed above for the problem of arranging the first n positive integers. While the problem is of a particular kind the approach is one that can be applied to a broader class of problems; for example, to assess the "randomness" of a mechanism that produces a sequence of 0's and 1's.

2. Specifying alternatives to randomness.

To specify a model for the problem of arranging the integers 1 to n requires specifying a distribution for the position of the number 1, n (conditional) distributions for the position of the number 2 (one for each possible position of the first number), n(n-1) distributions for the position of the number 3, (one for each ordered pair of positions selected initially), etc. It seems efficacious to reduce the size of this class of models. Many reductions are possible and the one considered here has no great virtue except that it provides quite a variety of alternatives to the random model - this class will be further reduced to a class of alternatives for each of which a trend from small numbers to large or large to small is regarded as being likely.

Let the random variable X_j denote the position in which the number j occurs, $j = 1,\ldots,n$. For nonnegative $n \times n$ matrix $P = (p_{ij})$ define

(1) $P(X_1 = i) = p_{i1} / \sum p_{i1}$ and

$$P(X_2 = i | X_1 = i') = p_{i2} / \sum_{\alpha \neq i'} p_{\alpha 2} ,$$

for $i = 1,\ldots,i'-1,\ i'+1,\ \ldots,\ n,$ so that at least two of the $p_{i2} > 0$ except that exactly one of p_{i2} can be positive so long as the corresponding $p_{i1} = 0$. In general, for $k = 2,\ldots,n,$ define

(2) $P(X_k = i | X_1 = i_1,\ldots,X_{k-1} = i_{k-1}) = p_{ik} / \sum p_{\alpha k}$,

for $i = 1,\ldots,n,\ i \neq i_1,\ldots,i \neq i_{k-1}$ and where the sum in the denominator excludes $\alpha = i_1,\ldots,\alpha = i_{k-1}$. Restrict consideration to matrices P for which the sum in (2) is positive for all k and all $(i_1,\ldots,i_{k-1})$. It would be sufficient, for example, to have at least k positive members of the k^{th} column of P, for $k = 1,\ldots,n$. The values p_{in} are immaterial, so long as none are 0, since the n^{th} number in the sequence is uniquely defined by the first $n-1$ numbers. (Actually, if the matrix P is such that the probability that i occurs before the n^{th} position is 1 then p_{in} can be 0, its value is completely immaterial.) Two different matrices can give rise to the same model. Obviously, any positive multiple of P is equivalent to P. The matrix with 1's on and above the main diagonal is equivalent to the identity matrix; under either, the probability of the arrangement $(1,2,\ldots,n)$ is 1.

Matrices which have identical (or proportional) columns deserve special consideration. They suggest a form of stationarity - provided positions i and i' have not been occupied by the numbers $1,2,\ldots,j-1$ the odds that j will occupy position i versus i' are the same for all j.

In the matrix which corresponds to the random model, call it P_0, every column is constant; that is, the rows are identical. We can as well take each member of P_0 to be 1.

Each member of the family of alternatives to P_0 considered here suggests a trend in the numbers selected. The smaller numbers are more likely to occur early in the sequence for some alternatives and late in the sequence for the others. These alternatives (or models) are indexed by the real number θ, which I call a "trend parameter". For all real θ every column of P_θ is identical;

$$p_{ij} = i^{-\theta} ,$$

for $i = 1,\ldots,n$ and all j. The family $\{P_\theta\}$ contains the random model P_0 as the special case $\theta = 0$. (For reasons of symmetry about $\theta = 0$ it is tempting to define p_{ij} to be $(n + 1 - i)^\theta$ for $\theta < 0$; I have resisted the temptation because the notation could obscure the main points of later arguments.) As an example consider $n = 5$ and $\theta = - 1$;

$$P_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 \end{pmatrix}.$$

Under $\underline{P}_{-1}$ the probability that the number 1 occurs in position i is $i/15$, $i = 1,2,3,4,5$; the $\underline{P}_{-1}$-probability that the number 2 occurs in position i given that the number 1 occurred in position 4, say, is $i/11$, $i = 1,2,3,5$; and so on.

3. The Likelihood Function of θ.

The likelihood function of θ for the arrangement $x = (i, i_2, \ldots, i_n)$ is

$$(3) \quad L_x(\theta) \propto \frac{i_1^{-\theta}}{\sum_{\alpha \neq i_1} \alpha^{-\theta}} \frac{i_2^{-\theta}}{\sum_{\substack{\alpha \neq i_1 \\ \alpha \neq i_2}} \alpha^{-\theta}} \cdots \frac{i_n^{-\theta}}{i_n^{-\theta}} .$$

Obviously, $L_x(0) \propto 1/n!$, a constant. For convenience we will normalize the function L_x by taking $L_x(0) = 1$; readers who prefer equality in (3) may regard our function L_x as the likelihood ratio $L_x(\theta)/L_x(0)$.

The likelihood function is reasonably well-behaved mathematically. L_x is continuous, and in fact differentiable, in θ for θ real and every arrangement x. L_x has a unique maximum; the modal value $\hat{\theta}$ is $+\infty$ for $x = (1,2,\ldots,n)$, $-\infty$ for $x = (n,n-1,\ldots,1)$, and finite for all other x's. The likelihood decreases monotonically to 0 away from $\hat{\theta}$ in either direction and it is bounded above by $n!$, a value approached only for the extreme arrangements $(1,2,\ldots,n)$ and $(n,n-1,\ldots,1)$, approached as $\theta \to +\infty$ and $\theta \to -\infty$, respectively.

Suppose $x_1,\ldots,x_m$ represent m independent arrangements produced using the same mechanism. The likelihood function of θ is now

$$L_{x_1,\ldots,x_m}(\theta) = \prod_{\beta=1}^{m} L_{x_\beta}(\theta) ,$$

which shares most of the characteristics of the individual L_{x_β}. In particular, the maximum likelihood value of θ, $\hat{\theta}$, is now finite unless the $x_1,\ldots,x_m$ are the same extreme arrangement.

A word of caution is in order concerning the class of models being considered here. Frequently in problems of statistical inference the observation actually obtained is the most probable observation under the maximum likelihood model. Since $\{P_\theta\}$ is a subclass of the class of models dictated by the structure of the experiment such is not the case in this problem. The maximal $P_{\hat{\theta}}$-probable sequence is $(1,2,\ldots,n)$ if $\hat{\theta} > 0$; $(n,n-1,\ldots,1)$ if $\hat{\theta} < 0$; and every sequence if $\hat{\theta} = 0$. If all models were considered and the independent arrangements $x_1,\ldots,x_m$ observed then the maximum likelihood model is the one which associates probability $1/m$ with each of $x_1,\ldots,x_m$; such a model has all the desirable asymptotic $(m \to \infty)$ properties but is hardly enlightening if m is small with respect to $(n!)$.

The assumption of independence is crucial. If it is incorrect the mechanism may be concluded random when it is far from random; it may even be deterministic! Suppose, for example, that $n = 2$ and that the mechanism produces first the sequence (1,2), then (2,1), and continues to alternate (1,2) and (2,1). For m even,

$$L_{x_1,\ldots,x_m}(\theta) = \left(\frac{2^{1+\theta/2}}{1 + 2^{\theta}}\right)^m ,$$

which is 1 for $\theta = 0$, symmetric about $\theta = 0$, and therefore, in view of previous discussion, $\hat{\theta} = 0$. For such a sequence of arrangements

$$L_{x_1,\ldots,x_m}(\theta) \to 0 \quad \text{as} \quad m \to \infty \quad \text{for all} \quad \theta: |\theta| > \varepsilon > 0$$

so the only reasonable conclusion (asymptotically) in the class of models considered is that the mechanism is random!

While arbitrary sample size m has been considered here, powerful inferences can be made for $m = 1$ if n is only moderately large. Roughly speaking, the information in a sample of size m is the order of nm. In the main application of the next section $m = 1$ and $m = 12$.

4. An Application: The 1970 U. S. Draft Lottery.

A contention made after the 1970 U. S. draft lottery was that, in my words, there was a trend in the resulting arrangement of the 366 days of the year. Regarding the days from January 1 to December 31 as numbered from 1 to 366, the contention was that the larger numbers were more likely to appear early in the arrangement; or, in the family $\{P_\theta\}$, that $\theta < 0$. The size of this example, $n = 366$, can be a hindrance to understanding the approach. Preliminarily, therefore, consider the resulting arrangement of the 12 months by rank of the average lottery number in the month. Indeed, it will be seen that the month is the appropriate unit for consideration! Using the natural ordering: January, February, and so on, the resulting arrangement was

$$x = (8,9,12,10,11,7,5,4,3,6,2,1) .$$

That is, January ranked eighth, February ninth, ..., and December first. The likelihood of θ is

$$(3) \qquad L_x(\theta) = \frac{8^{-\theta}}{\sum_1^{12} i^{-\theta}} \cdot \frac{9^{-\theta}}{\sum_{i \neq 8} i^{-\theta}} \cdot \frac{12^{-\theta}}{\sum_{\substack{i\neq 8\\ i\neq 9}} i^{-\theta}} \cdot \ldots \cdot \frac{1^{-\theta}}{1^{-\theta}} .$$

Several values of $L_x(\theta)$ are given in Table 1 and the function is graphed in Figure 1. Also given in Table 1 is the posterior probability of θ when its prior probability is 1/2 and the prior probability of the random model is 1/2. This probability is given by

$$(4) \qquad \Pr(\theta|x) = \left[1 + \frac{1-\Pr(\theta)}{L_x(\theta)\Pr(\theta)}\right]^{-1} = \left[1 + \frac{1}{L_x(\theta)}\right]^{-1} ,$$

and is the complement of the probability that the correct model is the random model: $\theta = 0$. If $L_x(\theta)$ is

[TABLE 1 ABOUT HERE]

[FIGURE 1 ABOUT HERE]

regarded as proportional to a posterior density (L_x can be normalized unless x is one of the two extreme arrangements), one corresponding to a uniform (improper) prior density on $(-\infty, \infty)$, then $P(\theta > 0|x) \approx 10^{-5}$. Therefore, assuming the true model is in $\{P_\theta\}$ and a prior density that is reasonably flat near 0, the trend parameter θ is very probably negative; that is, the trend is from large numbers to small, or, in terms of times of the year, from late to early.

If $\theta = -2.72$ then the vector of p_{i1}, normalized so that it is a probability vector, is

$$(.00031, .00204, .00615, .01345, .02468, .04053, .06164, .08864, .12211, .16263, .21076, .26704) .$$

For $\theta = -2.72$ January is 862 (or $12^{2.72}$) times as likely to be last (as it was in fact!) as first. Incidentally, for a uniform prior density on $(-\infty, \infty)$ the average value of this quantity is

$$E(12^{-\theta}|x) = +\infty .$$

The way in which the lottery was conducted makes clear the possibility of a trend. The following quote from (Fienberg 1971) which includes a description from the New York Times may even suggest a prior probability distribution for θ. Since an end of the box from which to pour was apparently

θ	-8	-6	-5	-4	-3	-2.7	-2	-1	0	+ 1
$L_x(\theta)$	1.8	41.9	161.7	473.9	893.0	913.9	703.6	118.3	1.0	.00007
Pr(θ'x)	.6429	.9767	.9939	.9979	.9989	.9989	.9986	.9916	--	.00007

TABLE 1: VALUES OF LIKELIHOODS IN (3) AND PROBABILITIES IN (4).

selected randomly it is regrettable that this prior distribution could not be assumed symmetric about 0; if the p_{ij} were defined for $\theta < 0$ to be $(n + 1 - i)^{\theta}$ (or vice versa) then the prior would be symmetric.

"A presidential proclamation, issued simultaneously with the executive order of 26 November 1969, stipulated that the 1970 lottery would be based on birthdays, and Selective Service officials devised the actual method of drawing the dates. Although an official detailed description of the actual procedures used is not available, Captain William Pascoe, chief of public information for the Selective Service System and the man in charge of the lottery, has informed me that the following account which appeared in the New York Times is basically correct.

'Over the weekend before the December 1st drawing, Captain Pascoe and Col. Charles R. Fox, under the watch of John H. Adams, an editor of U. S. News and World Report, set up the lottery.

They started out with 366 cylindrical capsules, one and a half inches long and one inch in diameter. The caps at the ends were round.

The men counted out 31 capsules and inserted in them slips of paper with the January dates. The January capsules were then placed in a large, square wooden box and pushed to one side with a cardboard divider, leaving part of the box empty.

The 29 February capsules were then poured into the empty portion of the box, counted again, and then scraped with the divider into the January capsules. Thus, according to Captain Pascoe, the January and February capsules were thoroughly mixed.

The same process was followed with each subsequent month, counting the capsules into the empty side of the box and then pushing them with the divider into the capsules of the previous months.

Thus, the January capsules were mixed with the other capsules 11 times, the February capsules 10 times, [11 times?] and so on with the November capsules intermingled with others only twice and the December ones only once.

The box was then shut, and Colonel Fox shook it several times. He then carried it up three flights of stairs, a process that Captain Pascoe says further mixed the capsules.

The box was carried down the three flights shortly before the drawing began. In public view, the capsules were poured from the black box into the two-foot deep bowl.

Captain Pascoe said he did not know which end of the box he poured from. If he poured from the end where the capsules with the early months had been repeatedly shoved, these capsules might have fallen to the bottom of the bowl. Conversely, if he poured from the other end, the later months could have fallen to the bottom. This assumes that the shoving and shaking procedure did not adequately mix the capsules.

Once in the bowl, the capsules were not stirred... The persons who drew the capsules last month generally picked ones from the top, although once in a while they would reach their hand to the middle or the bottom of the bowl.'

Once again the question of inadequate mixing of capsules must be raised. From the above description one might expect that dates late in the year would tend to be drawn early, and dates early in the year would tend to come up late in the drawing (or vice versa if in fact the box had been turned around)."

Considering now the 366 numbers in the actual lottery (for the arrangement x see (Fienberg 1971)) the likelihood is maximized near $\theta = -.25$, and $L_x(-.25) = 89{,}700$. In view of the way in which the capsules were mixed, however, the analysis by months ($n = 12$) seems more reasonable (if you like, the corresponding class of models has higher prior probability). As a check, I have analyzed the relative ranks of the days in each of the months and found that there does not seem to be a trend in any of the twelve months.

5. A Response to I. D. Hill.

In a recent article (1974), I. D. Hill challenges the "likelihood" and Bayesian schools of inference to come up with an approach to a question relating to the degree of involvement of chance, as against skill, in football games. He reformulated the question by comparing "the final league tables with expert forecasts made before the start of the season".

Consider Hill's Table 1, Football League Division 1, 1971-72. Label the teams forecasted to finish from first (Tottenham) to twenty-second and last (Crystal Palace) as numbers 1 to 22. The actual ordering determined by final standings was

$$x = (6,2,7,3,5,9,15,8,1,4,19,16,11,18,17,10,14,21,12,13,22,20) .$$

The question is: Does this sequence exhibit a trend? The conclusion that $\theta \neq 0$ (presumably positive) implies that more than chance is involved in forecasting the final standings, and therefore in the game of football itself. The likelihood of θ for x is presented in Table 2, along with posterior probabilities assuming $Pr(\theta) = Pr(0) = 1/2$, and Figure 2.

[TABLE 2 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

I have found (numerically) the maximum likelihood estimates of θ for each of Hill's Tables 1-6, corresponding to 1971-72 Football League Divisions 1 to 4 and Scottish League Divisions 1 and 2, and present these in Table 3. The value of n is the number of teams in the division and

θ	-.25	0	.25	.50	.75	.90	1.00	1.25	1.50	1.75	2.00
$L_x(\theta)$	.085	1	7.92	37.17	89.67	105.08	99.18	48.49	11.14	1.37	.104
$Pr(\theta \| x)$	.0783	--	.8879	.9738	.9890	.9906	.9900	.9798	.9176	.5781	.0942

TABLE 2: VALUES OF LIKELIHOODS AND POSTERIOR PROBABILITIES FOR HILL'S FOOTBALL LEAGUE DIVISION 1 DATA.

	FLD 1	FLD 2	FLD 3	FLD 4	SLD 1	SLD 2
n	22	22	24	24	18	19
$\hat{\theta}$	.90	.50	1.32	.38	2.43	.87
$L_x(\hat{\theta})$	105.08	6.03	965.57	2.58	9096.45	15.27
Kendall's τ	.5238	.3593	.4493	.1304	.5686	.3801

TABLE 3: MAXIMUM LIKELIHOOD ESTIMATES FOR HILL'S TABLES 1 - 6.

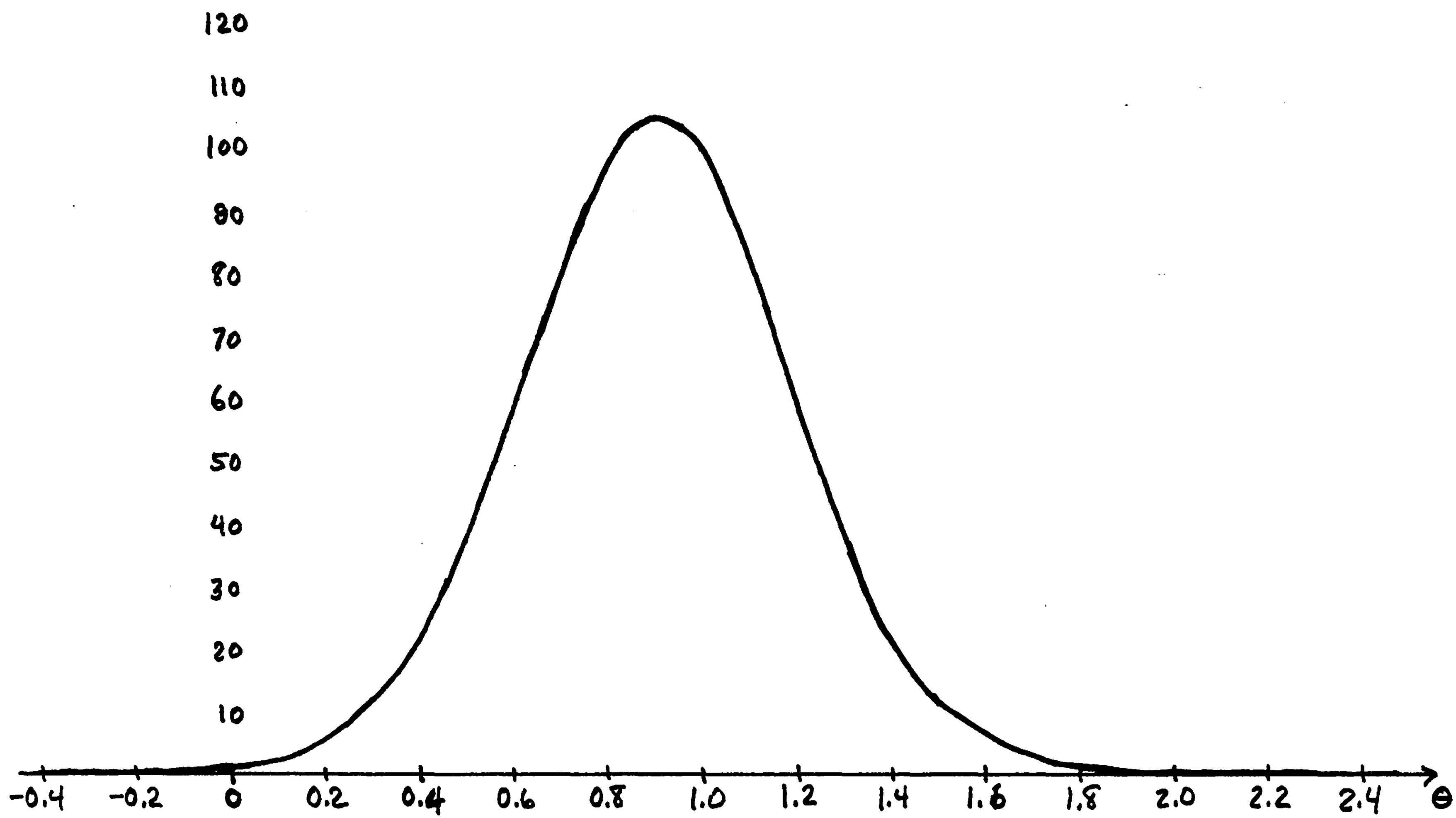

FIGURE 2: $L_x(\theta)$ FOR $x=(6,2,7,3,5,9,15,8,1,4,19,16,11,18,17,10,14,21,12,13,22,20)$

Kendall's τ is Kendall's rank correlation coefficient as computed by Hill. It is conceivable that "skill" was as much involved in Football League Division 4 as in Division 1, for example, but the forecasters were not as familiar with Division 4 as they were with Division 1.

Different but related models can be expected to yield higher likelihoods. I have obtained more than double the maximum of the likelihood shown in Table 3 by simply relabelling 22 to 1 instead of 1 to 22, or equivalently, by defining $p_{ij} = (n + 1 - i)^{\theta}$ with the original labelling.

When seeing Table 3 classically minded statisticians (and many Bayesians too!) will ask for the sampling distribution of $L_x(\hat{\theta})$ when in fact $\theta = 0$; that is, when the arranging mechanism is actually random. Asymptotically (as $mn \to \infty$) $L_x(\hat{\theta})$ tends to 1 with probability 1 when $\theta = 0$. I have not done extensive simulation for any finite pair (m,n) but I have done enough for $m=1$ and $18 \leq n \leq 24$ to suggest that 2.58 for FLD 4 is within the range of reasonable values. The numbers 6.03 and 15.27 for FLD 2 and SLD 2 are quite large, perhaps near the 5% level and 1% level, respectively. The remainder of the $L_x(\hat{\theta})$ in Table 3 are "very significant" to say the least. The number 9096.45 for Scottish League Division 1 is remarkably large; the corresponding arrangement is

$$x = (1,3,2,8,15,4,9,6,5,12,18,7,10,14,13,11,17,16) .$$

While I have not determined the distribution of $L_x(\hat{\theta})$ when $\theta = 0$ it is clear that it has a heavy tail. While most of the probability (for moderate or large n) is concentrated between 1 and 3 the expected value of $L_x(\hat{\theta})$ is greater than 3 (for $n \geq 4$) since

$\Pr(L_x(\hat{\theta}) = n! | \theta = 0) = (2/n!)$

6. Another Application.

As 39 students in an elementary statistics course handed in their final examination I kept them in order, the latest handed in on top and the earliest on the bottom. They were handed in over a period of 15 minutes with about half handed in during the last 2 minutes. The question is: What sort of trend, if any, is present in the grades received? If there is no trend then the ranks of the grades (1 for the highest grade, etc.) will be arranged randomly.

The actual arrangement was

$$x = (29,11,24,27,39,32,30,4,22,24,20,37,7,10,21,35,19,13,34,23,9,25,26,$$
$$18,12,2,17,36,33,8,6,38,28,5,3,16,31,15,1)\ .$$

The graph of $L_x(\theta$ is shown in Figure 3.

[FIGURE 3 ABOUT HERE]

For this arrangement $\hat{\theta} = -.39$ and $L_x(-.39) = 11.142$. The suggested negative trend means that higher grades are more likely later in the sequence; that is, handed in later. If $\theta = -.39$ then, for example, the probability that the top examination, the latest handed in, is the highest grade is .00898 and the lowest grade is .03747.

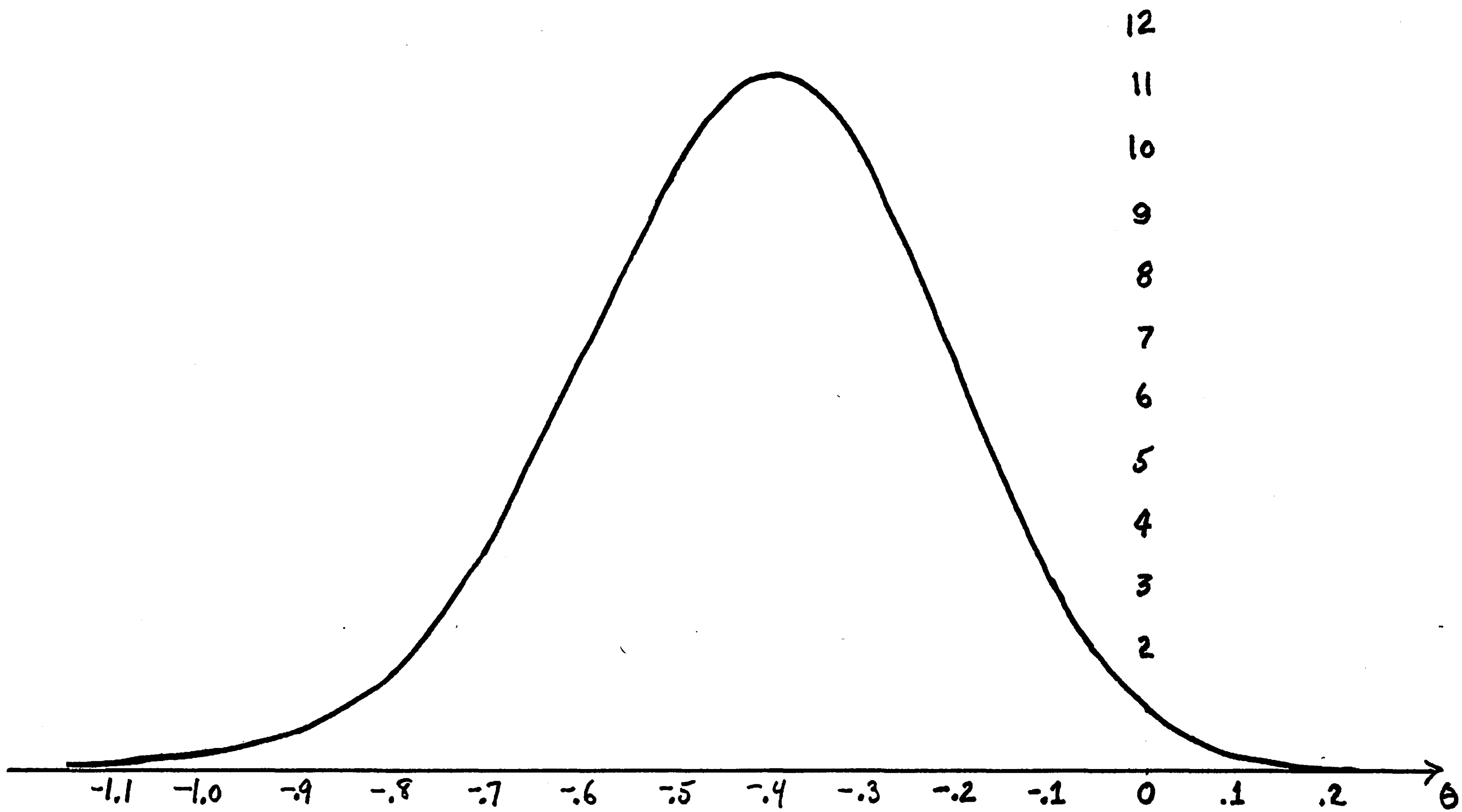

FIGURE 3

7. Concluding Remarks.

In my view the only way that a stochastic model (or class of models) can be indicated or counterindicated is to imbed it in a larger class of models. Its posterior probability (given this class) can then be calculated using Bayes' theorem. The process is not cut and dried: there is no number above which this probability must fall to indicate the model as the "true" one or a number below which counterindicates the model. Indeed the true model is almost never contained in a particular class of models!

For the problem of determining whether the mechanism that arranges the first n integers in a sequence is random, a larger class of models each member of which suggests a trend in the numbers from large to small or small to large has been considered here. This class is conveniently indexed by a real number θ, called a trend parameter.

There are, of course, many ways to hypothesize a trend. One that suggests itself is to define

$$p_{ij} = \begin{cases} 1 - \gamma i\,, & \text{if } \gamma \le 0 \\ 1 + \gamma(n+1-i), & \text{if } \gamma \ge 0 \end{cases}$$

for $i,j = 1,\ldots,n$. γ then could be called a trend parameter and $\gamma = 0$ corresponds to the random model. One annoyance caused by this definition is that the likelihood function of γ tends to a positive constant as $\gamma \to \pm\infty$ for any arrangement. Frequently the likelihood function is monotonic; for example, in the case of the ranks of the months in the 1970 draft lottery the likelihood is strictly decreasing.

Acknowledgment:

I wish to thank my friend, Christopher Bingham, for calling my attention to the Hill paper.

References

Fienberg, Stephen E. (1971). Science, 171, 255-261.

Hill, I. D. (1974). Association football and statistical inferences, Appl. Statist., 23, 203-208.

New York Times, January 4, 1970, p. 66.