UCD

83-3

UNIV. CALIF.

UCD Department of Agricultural Economics

WORKING PAPER SERIES

University of California, Davis
Department of Agricultural Economics

A RELATIONSHIP BETWEEN HYPOTHESES TESTING AND LINEAR PROGRAMMING: A PEDAGOGICAL NOTE

by

Richard Green

Working Paper No. 83-3

A RELATIONSHIP BETWEEN HYPOTHESES TESTING AND LINEAR PROGRAMMING: A PEDAGOGICAL NOTE

I. INTRODUCTION

The purpose of this note is to show that linear programming is related very closely to the famous statistical theorem of Neyman and Pearson. The theorem, sometimes referred to as the fundamental lemma of Neyman and Pearson, states that the likelihood ratio test is a most uniformly powerful test of size α for testing a simple hypothesis against a simple alternative. Although most agricultural economists are familiar with both linear programming and the likelihood ratio testing procedure, it is not commonly known that the two methodological approaches are conceptually based on the same principles.

After a more rigorous treatment of the Neyman-Pearson lemma, a typical linear programming problem is presented which illustrates that the ideas contained in the Neyman-Pearson can be used to obtain the same optimal solution to the problem. The paper concludes by noting the close resemblance between the statistical concepts developed by Neyman and Pearson and the approaches often times taken by applied economic researchers.

II. NEYMAN-PEARSON LEMMA

In 1933 Neyman and Pearson published a seminal paper on the theory of hypothesis testing. In the paper they formulated the hypothesis testing problem in terms of two types of errors and introduced the likelihood ratio criterion as a general method of test construction. A precise formulation of the Neyman-Pearson lemma is given as follows (Lehmann, p. 65):

Theorem: Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1, respectively, with respect to a measure μ. For testing $H{:}p0$ against the alternative $K{:}p1$ there exists a test

ϕ and a constant k such that

(1) $E_0\ \phi(X) = \alpha$

and

(2) $\phi(x) = \begin{cases} 1 \text{ when } p_1(x) > k\ p_0(x) \\ 0 \text{ when } p_1(x) < k\ p_0(x). \end{cases}$

If a test satisfies (1) and (2) for some k, then it is most powerful for testing p_0 against p_1 at level α. If ϕ is most powerful at level α for testing p_0 against p_1, then for some k it satisfies (2) a.e., μ. It also satisfies (1) unless there exists a test of size $< \alpha$ and with power 1.

To better understand the theoretical construction of the lemma suppose there exists two discrete distributions P_0 and P_1. Let P_0 be associated with the simple null hypothesis and let P_1 hold under the simple alternative. Considering only nonrandomized tests, the optimal testing procedure is given by [Lehmann, p. 64] the critical region C which satisfies

$$\sum_{x \varepsilon C} P_0(x) \leq \alpha$$

and

$$\sum_{x \varepsilon C} P_1(x) = \text{maximum}.$$

That is, the sample points contained in C under P_0 should have a total value, $\Sigma P_0(x)$, of less than or equal to α, the size of the test. Simultaneously, the value of the sample points in C assuming that P_1 is true should be as large as possible. The latter statement is just an expression of maximizing the power of the test. Given this interpretation of the hypothesis testing formulation, the method to select the optimal test or critical region consists of ranking the values of the ratios $P_1(x)/P_0(x)$ starting with the largest value.

Then include as many of the sample points in C without violating the condition, $\sum_{x \in C} P_0(x) \leq \alpha$.

The following simple statistical example illustrates the procedure in determining the optimal test. Suppose we are interested in testing $H_0\colon \pi_0 = 0.50$ versus $H_1\colon \pi_1 = 0.40$ where the underlying probability distributions are both binomial. Thus, we are interested in testing whether the population proportion is 0.50 or 0.40. If our sample size were ten, then we would have the probability values given in Table 1.

The sample points x = 0, 1, 2, . . .(zero, one, . . ."success") yield the highest values for the ratio $P_1(x)/P_0(x)$. Thus, given that $\alpha = 0.02$ say, the optimal critical region would consist of the sample points x = 0 and x = 1. These points give rise to $\sum_{x=0}^{1} P_1(x) = 0.0108 \leq 0.02$, hence satisfying the condition that the size of the test not exceed 0.02. Also the value of $\sum_{x=0}^{1} P_1(x)$ is maximized for $\alpha \leq 0.02$.

A minor technical problem still exists. To achieve an exact value α, randomized tests can be performed. These tests will not be considered here since they would add little to the main conceptual point of the paper. See, e.g., Mood *et al.*, for a discussion of randomized tests.

III. LINEAR PROGRAMMING PROBLEM

Now consider the following problem (refer to Table 2 for the data):[1] I have a budget of \$186 and wish to get as many pages as possible for this amount. What books should I purchase? When I was first confronted with this problem, I immediately realized that a linear (integer) programming method would provide the optimal solution. However, it was not obvious that the

Table 1. Values of $P(x)$ for Binomial Probability Distribution with $n = 10$

x	$P_1(x)$	$P_0(x)$	$P_1(x)/P_0(x)$
0	.0060	.0010	6.00
1	.0403	.0098	4.11
2	.1209	.0439	2.75
3	.2150	.1172	1.83
4	.2508	.2051	1.22
5	.2007	.2461	0.81
6	.1115	.2051	0.54
7	.0425	.1172	0.36
8	.0106	.0439	0.24
9	.0016	.0098	0.16
10	.0001	.0010	0.10

Source: Tables of the Binomial Probability Distribution, Applied Mathematics Series, (U.S. Department of Commerce, 1950).

solution procedure also illustrated a famous statistical theorem of Neyman and Pearson. The fixed budget of $186 corresponds to a fixed α in a hypothesis testing problem. Next form the ratios involving the number of pages per dollar and rank then starting with the highest value (book #12); see Table 2. Then include as many sample points (books) as possible without violating the budget of $186. This results in purchasing books numbered, 12, 10, and 3 with a total value of $177.95. The discrete nature of the problem precludes an exact solution. These sample points (books #12, 10 and 3) correspond to those points which give the most power for a type I error (budget of $186).

Thus, what appears and is a linear (integer) programming problem turns out to be solvable by employing the Neyman-Pearson lemma. The difficulty of obtaining an exact α can be overcome by resorting to randomized tests (Lehmann).

IV. CONCLUSIONS

Agricultural economists are well trained in quantitative techniques and in particular with linear programming and hypotheses testing. However, few, the author believes, realize the close connection between linear programming and the famous theorem developed by Neyman and Pearson. In fact, the primary idea underlying the Neyman-Pearson lemma is that of attempting to maximize the "value per dollar" given a "fixed budget" (Lehmann). This situation occurs often in economics, for example, in the theory of consumer behavior. Consequently, many of the optimization models of economics illustrate a statistical concept which serves as a basic foundation for the theory of hypothesis testing.

Table 2. Data on Books, Number of Pages, Prices and Ratio Values

Book No.	No. of Pages $P_1(x)$	Price of Book $P_0(x)$	Sample Pt. (Book No.)	$P_1(x)/P_0(x)$[1]
1	400	$ 69.00	12	15.98
2	237	34.50	10	10.97
3	474	45.00	3	10.53
4	606	69.50	20	9.23
5	600	130.00	17	9.14
6	760	160.00	4	8.72
7	272	69.50	9	8.33
8	148	35.00	18	8.22
9	400	48.00	13	7.63
10	384	35.00	16	6.89
11	266	69.00	21	6.87
12	175	10.95	2	6.87
13	530	69.50	1	5.80
14	392	112.00	19	5.79
15	400	74.00	15	5.41
16	272	39.50	6	4.75
17	896	98.00	5	4.61
18	448	54.50	22	4.54
19	420	72.50	8	4.23
20	803	87.00	7	3.91
21	268	39.00	11	3.85
22	304	67.00	14	3.50
23	270	82.00	23	3.29

[1]Ratios are listed in descending order.

FOOTNOTES

1. This problem was given on a midterm by Professor W. Thompson in a statistics course at the University of Missouri while the author was auditing the course while on sabbatic leave from the University of California and a visiting associate professor of the University of Missouri.

REFERENCES

Lehmann, E. L., *Testing Statistical Hypotheses*. New York: John Wiley and Sons, 1959.

Mood, A., F. Graybill and D. Boes, *Introduction to the Theory of Statistics*, 3rd ed., New York: McGraw-Hill Book Company, 1974.

Neyman, J. and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Phil. Trans. Roy. Soc., Ser. A.*, Volume 231, 1933, pp. 289-337.