1986

Improving upon the Neyman-Person Approach to Testing Hypotheses

Noel Cressie

Peter B. Morgan

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

Improving Upon The Neyman-Pearson Approach To Testing Hypotheses

Noel Cressie
Peter B. Morgan

TECHNICAL REPORT NO. 13
JULY 1986

Centre For Decision Sciences And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

2813

ISSN 0831-4454
ISBN 0-7714-0761-0

IMPROVING UPON THE NEYMAN-PEARSON APPROACH TO TESTING HYPOTHESES

NOEL CRESSIE[1] and PETER B. MORGAN[2]

Preprint Number 86-19

March 1986

[1]Department of Statistics
Iowa State University
Ames, IA 50011
U.S.A.

[2]Department of Economics
University of Western Ontario
London, Ontario N6A 5C2
Canada

IMPROVING UPON THE NEYMAN-PEARSON APPROACH TO TESTING HYPOTHESES

by

Noel Cressie and Peter B. Morgan
Iowa State University and University of Western Ontario

SUMMARY

In this paper we place the Neyman-Pearson testing procedure into a decision theoretic context, where collecting observations incurs a given cost, and making right/wrong decisions yield various given payoffs. We show that in general the Neyman-Pearson test generally does not maximize the expected payoff net of costs, and we give the procedure which does. Furthermore, we connect this optimality notion with that of maximizing power subject to bounding the size.

Some key words: cost function, decision theory, maximizing expected net gain, payoffs, power, sequential test, size.

1. Introduction

Suppose a decision must be made about which of the two elements of the parameter space of $\Omega = \{\omega_0, \omega_1\}$, is governing the distribution function $F(\underset{\sim}{x};\omega)$ of data $\underset{\sim}{x}$. Let the distributions be absolutely continuous with respect to a dominating measure and write $f(\underset{\sim}{x};\omega)$ as the Radon-Nikodym derivative. For continuous data, the dominating measure is the Lebesgue measure and f becomes the density; for discrete data, the counting measure is used and f becomes the probability mass function.

The Neyman-Pearson lemma (Neyman and Pearson, 1933) states that the "best" decision about the truth of the hypotheses, $H_0: \omega = \omega_0$, versus $H_1: \omega = \omega_1$, is given by the decision function:

$$\Phi^*(\underset{\sim}{x}) = \begin{cases} 1, & \text{if } \dfrac{f(\underset{\sim}{x};\omega_0)}{f(\underset{\sim}{x};\omega_1)} < k \\ \gamma, & \text{if } \dfrac{f(\underset{\sim}{x};\omega_0)}{f(\underset{\sim}{x};\omega_1)} = k \\ 0, & \text{if } \dfrac{f(\underset{\sim}{x};\omega_0)}{f(\underset{\sim}{x};\omega_1)} > k \end{cases}, \qquad (1.1)$$

where $\Phi^*(\underset{\sim}{x})$ is interpreted as the probability of rejecting H_0; rejection of one of the hypotheses is equivalent to acceptance of the other. "Best" here means that for any other decision function Φ, $0 \le \Phi \le 1$, satisfying $E_{\omega_0}(\Phi(\underset{\sim}{x})) \le E_{\omega_0}(\Phi^*(\underset{\sim}{x})) \equiv \alpha$, then

$$E_{\omega_1}(\Phi(\underset{\sim}{x})) \le E_{\omega_1}(\Phi^*(\underset{\sim}{x})). \qquad (1.2)$$

The quantities in (1.2) are called the power of their respective decision functions, and in statistical parlance Φ^* is said to maximize the power amongst all tests Φ of size at most α. In the theory of hypothesis testing (Lehmann, 1959, p. 61), the main problem is to find k and γ in (1.1) such that the size α is fixed at say .05 or .01, etc.; the resulting test Φ^* has maximum power for its size.

But this approach to hypothesis testing is incomplete in that it offers no guidance as to how to actually do the sampling. While having more observations typically increases the power, there are usually practical limitations to the amount of data that can be gathered. In particular, there is a cost associated with gathering data $\underset{\sim}{x}$, and this cost is not addressed by the Neyman-Pearson lemma.

We shall show in the following sections that a more sensible approach to testing recognizes the payoff and cost functions faced by the decisionmaker. Doing so allows computation of an optimal sampling strategy from which the optimal size, $E_{\omega_0}(\Phi^*(\underset{\sim}{x}))$, and optimal power $E_{\omega_1}(\Phi^*(\underset{\sim}{x}))$ are calculable directly from the payoff and cost functions and p_1, the prior probability that $\omega = \omega_0$. In this way the customary hypothesis testing procedure is made more informative by adding decision-theoretic concepts to the traditional method. Section 2 sets up the formalities needed for Section 3, which explains how to improve the Neyman-Pearson approach. Section 4 explains why the size, power, and sample size of the optimal procedure are functions of the payoffs, cost, and prior probability; two examples are given. Section 5 discusses implications for the Neyman-Pearson test. For example, fixing the size of the test and the number in the sample is shown to put restrictions on p_1 which may exclude the noninformative prior $p_1 = 1/2$.

2. Testing two simple hypotheses

Consider for the moment a more general problem, sequential in nature, which allows the possibility of a decision about ω at any of $t = 1,2,\ldots,T$ discrete decision points, $1 \le T \le \infty$; T is called the truncation point. In this paper we will be concerned with the special case $T = 2$; see Cressie and Morgan (1986) for the general case, where they show the optimal decision rule to be a variable-sample-size-sequential probability ratio test, called the VPRT. The sequential probability ratio test, or SPRT, is a constrained version of the VPRT, whose optimality properties have been studied extensively since the seminal papers of Wald and Wolfowitz (1948), and Arrow, Blackwell, and Girshick (1949).

For $n \ge 0$, let $c(n)$ be the cost incurred at any decision point by asking for n observations. If $n_t \ge 1$ observations $(x_{t,1},\ldots,x_{t,n_t}) \equiv \underset{\sim}{x}_t$ are asked for at time t, then these are received at time $t + 1$. Assume that $c(n)$ is strictly increasing with n. Let $c_0 \equiv c(0)$ denote the "overhead" or "fixed cost" of sampling which is incurred by waiting the period until the next decision point, without collecting any observations.

In the remainder of this section we describe the meaning of the term "decision rule." A formal and detailed development is given in Cressie and Morgan (1986); here we summarize that development. A decision rule has three components: a terminal decision rule, a sample-size rule, and a stopping rule. Each of these is described in turn.

A terminal decision rule δ is a sequence $\{\delta_t;\ t = 1,\ldots,T\}$. The element δ_t is a function of all available data thus far, viz.

$$\underset{\sim}{y}_t = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_{t-1}), \text{ and}$$

$$\delta_t(\underset{\sim}{y}_t) \equiv \begin{cases} 0, & \text{if } \omega_0 \text{ is chosen} \\ 1, & \text{if } \omega_1 \text{ is chosen}; \quad t = 2, \ldots, T. \end{cases} \tag{2.1}$$

At any decision point t, a choice is made between continuing to take more observations, and stopping to make a terminal decision. The "payoffs" or "rewards" for these terminal decisions, are defined by the payoff function:

$$U(\omega, \delta_t) = \begin{cases} u_{00}, & \text{if } \delta_t = 0 \text{ and } \omega = \omega_0 \\ u_{01}, & \text{if } \delta_t = 0 \text{ and } \omega = \omega_1 \\ u_{10}, & \text{if } \delta_t = 1 \text{ and } \omega = \omega_0 \\ u_{11}, & \text{if } \delta_t = 1 \text{ and } \omega = \omega_1 . \end{cases} \tag{2.2}$$

The units are the same as those for the cost function: pounds, dollars, francs or whatever numeraire is meaningful to the decisionmaker. We have decided to take the more optimistic viewpoint of looking at payoffs rather than losses; those familiar with loss functions should have no trouble thinking in terms of payoff functions. Typically, $u_{00} > u_{10}$ and $u_{11} > u_{01}$ since it is usual to reward correct decisions more than incorrect ones.

A sample-size rule ν is a sequence $\{\nu_t;\ t = 1, \ldots, T\}$. This component of a sequential decision procedure is usually not mentioned in the sequential analysis literature, since therein one-at-a-time sampling is most often imposed. The element $\nu_t \geq 0$ is a function of all available data thus far, viz. $\underset{\sim}{y}_t = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_{t-1})$, and tells the decisionmaker how many observations should be asked for at time t, to be received at time t+1, if sampling is to continue. It is this generality which allows us to ask for and find an optimal ν, and hence to develop procedures which dominate the more classical ones.

A stopping rule S is a sequence $\{S_t;\ t = 1,\ldots,T\}$. Each element S_t is a function of all the data available at t, viz. $\underset{\sim}{y}_t$, with

$$S_t(\underset{\sim}{y}_t) \equiv \begin{cases} 0, & \text{if sampling continues at } t \\ 1, & \text{if sampling stops at } t\ , \end{cases} \qquad (2.3)$$

for $t = 2,\ldots,T-1$; and $S_T(\cdot) \equiv 1$.

The special cases of δ_1, ν_1, S_1, which are not based on data, are determined from the prior probability $p_1 = pr(\omega = \omega_0)$. If $\nu_{t-1} > 0$, this prior is updated to give a posterior p_t at time t:

$$p_t = \frac{f(\underset{\sim}{x}_{t-1};\omega_0)p_{t-1}}{f(\underset{\sim}{x}_{t-1};\omega_0)p_{t-1} + f(\underset{\sim}{x}_{t-1};\omega_1)(1-p_{t-1})}\ ;\ t = 2,\ldots,T; \qquad (2.4)$$

if $\nu_{t-1} = 0$, $p_t = p_{t-1}$.

Finally, a decision rule d is defined to be the ordered triple $(S,\ \nu,\ \delta)$. Our goal is to find an "optimal" decision rule d*. Here optimal means maximizing the expected payoff minus cost, or what we call the <u>expected net gain</u>. When the problem is formulated in terms of losses rather than payoffs, d* is simply the Bayes procedure minimizing the Bayes risk.

In the next section we present the various expected net gains for continuing versus those for stopping, for the special case $T = 2$. The general case is more complicated, and details can be found in Cressie and Morgan (1986); the optimal strategy is the intuitively obvious one of choosing from among the three actions: stop and select ω_0, continue collecting observations, stop and select ω_1; that action which gives the highest expected net gain. This strategy for T = 2 leads to an improvement on the Neyman-Pearson approach to testing hypotheses. Not only does the approach answer the design question of how many observations to take, but as well provides insight into the true nature of the size and power of a test procedure.

3. Maximizing expected net gain

At the initial decision point t = 1, a choice must be made between selecting ω_0, selecting ω_1, and collecting observations. The expected net gain from selecting ω_0 is

$$u_{00}p_1 + u_{01}(1-p_1). \tag{3.1}$$

The expected net gain from selecting ω_1 is

$$u_{10}p_1 + u_{11}(1-p_1). \tag{3.2}$$

Hence a terminal decision that maximizes the expected net gain from not collecting observations, is

$$\delta_1^* = \begin{cases} 0, & \text{if } u_{00}p_1 + u_{01}(1-p_1) \geq u_{10}p_1 + u_{11}(1-p_1) \\ 1, & \text{otherwise .} \end{cases} \tag{3.3}$$

The action of collecting observations $\underset{\sim}{x}_1$ and subsequently making an inference about ω at t = 2 = T, has a maximal expected net gain

$$-c(n_1^*) + E_{\underset{\sim}{x}_1}[\max\{u_{00}p_2+u_{01}(1-p_2),\ u_{10}p_2+u_{11}(1-p_2)\}|p_1,n_1^*], \tag{3.4}$$

where n_1^* is

$$\mathrm{argmax}[-c(n_1) + E_{\underset{\sim}{x}_1}\{\max(u_{00}p_2+u_{01}(1-p_2),\ u_{10}p_2+u_{11}(1-p_2))|p_1,n_1\};\ n_1 \geq 0], \tag{3.5}$$

p_2 is the posterior probability of ω_0 given by (2.4), and $E_{\underset{\sim}{x}_1}$ denotes the expectation with respect to the distribution function

$$F_1(\underset{\sim}{x}) \equiv p_1F(\underset{\sim}{x};\omega_0) + (1-p_1)F(\underset{\sim}{x};\omega_1) . \tag{3.6}$$

In words, n_1^* is chosen to maximize the expected net gain of collecting observations. Define this optimal sample size rule,

$$\nu_1^*(p_1) \equiv n_1^*.$$

If $t = 2 = T$ is reached, a choice between selecting ω_0 and selecting ω_1 must be made. A terminal decision rule which maximizes expected net gain at $t = 2$, is

$$\delta_2^* = \begin{cases} 0, & \text{if } u_{00}p_2+u_{01}(1-p_2) \geq u_{10}p_2+u_{11}(1-p_2) \\ 1, & \text{otherwise} \end{cases}$$

$$= \begin{cases} 0, & \text{if } \dfrac{f(\underset{\sim}{x}_1;\omega_0)}{f(\underset{\sim}{x}_1;\omega_1)} \geq \dfrac{(1-p_1)(u_{11}-u_{01})}{p_1(u_{00}-u_{10})} \\ \\ 1, & \text{if } \dfrac{f(\underset{\sim}{x}_1;\omega_0)}{f(\underset{\sim}{x}_1;\omega_1)} < \dfrac{(1-p_1)(u_{11}-u_{01})}{p_1(u_{00}-u_{10})} , \end{cases} \tag{3.7}$$

which is exactly of the form of the Neyman-Pearson test function Φ^* given by (1.1) with $\gamma = 0$. It is obvious that the expected net gain remains unchanged when using a δ_2 that randomizes over selection of ω_0 or ω_1 when

$$\frac{f(\underset{\sim}{x}_1;\omega_0)}{f(\underset{\sim}{x}_1;\omega_1)} = \frac{1-p_1(u_{11}-u_{01})}{p_1(u_{00}-u_{10})} .$$

Hence the randomized version (1.1) also maximizes expected net gain at $t = 2$.

The expected net gain of the entire two period decision procedure is

$$\max\{(3.1), (3.2), (3.4)\}.$$

So a stopping rule that gives this expected net gain is

$$S_1^* = \begin{cases} 0 \text{ (i.e. observations are collected)}, & \text{if } (3.4) > \max\{(3.1),(3.2)\} \\ 1 \text{ (i.e. select } \omega_0 \text{ or } \omega_1) & , \text{ if } (3.4) \le \max\{(3.1),(3.2)\} \end{cases} \tag{3.8}$$

Theorem 3.1 of Cressie and Morgan (1986) shows that there exist calculable values p_L and p_U, $0 \le p_L \le p_U \le 1$, such that the optimal decision rule can be written in the simple form:

$$\begin{aligned} S_1^* &= 1 \text{ and } \delta_1^* = 1 , && \text{if } p_1 \in [0,p_L] \\ S_1^* &= 0 \text{ and } \nu_1^* = n_1^*, && \text{if } p_1 \in (p_L,p_U) \\ S_1^* &= 1 \text{ and } \delta_1^* = 0 , && \text{if } p_1 \in [p_U,1] . \end{aligned} \tag{3.9}$$

The size and power of the optimal test procedure (3.9) and (3.7) are

$$\alpha^* = \begin{cases} 1 & ; \ p_1 \in [0,p_L] \\ \mathrm{pr}\left(\frac{f(\underset{\sim}{x}_1;\omega_0)}{f(\underset{\sim}{x}_1;\omega_1)} < \frac{(1-p_1)(u_{11}-u_{01})}{p_1(u_{00}-u_{10})} \,\middle|\, \omega = \omega_0\right) & ; \ p_1 \in (p_L,p_U) \\ 0 & ; \ p_1 \in [p_U,1] , \end{cases} \tag{3.10}$$

$$\pi^* = \begin{cases} 1 & ; \ p_1 \in [0,p_L] \\ \mathrm{pr}\left(\frac{f(\underset{\sim}{x}_1;\omega_0)}{f(\underset{\sim}{x}_1;\omega_1)} < \frac{(1-p_1)(u_{11}-u_{01})}{p_1(u_{00}-u_{10})} \,\middle|\, \omega = \omega_1\right) & ; \ p_1 \in (p_L,p_U) \\ 0 & ; \ p_1 \in [p_U,1] . \end{cases} \tag{3.11}$$

Since p_L, p_U are functions of u_{00}, u_{01}, u_{10}, u_{11}, $c(\cdot)$, and p_1, it follows from (3.10) and (3.11) that the size and power of the optimal test procedure are also functions of these quantities.

Also observe that the optimal sample size n_1^* is similarly a function of the payoffs, the cost and the prior; that is, any choice of n_1 which does not respect this functional relationship will result in a testing procedure of lower expected net gain, and a different size and power. For example, the Neyman-Pearson testing procedure which uses a fixed sample size does not in general maximize the expected payoff net of cost. In the next section we connect the optimality notion of maximizing expected net gain, to the more familiar one of maximizing power subject to a bound on the size.

4. The computation of size, power, and optimal sample size

From (3.5), it is clear that the optimal sample size $n_1^*(u_{00}, u_{01}, u_{10}, u_{11}, c(\cdot), p_1)$, is a function of payoffs, cost, and prior. Similarly from (3.10) and (3.11), size and power can be written as $\alpha^*(u_{00}, u_{01}, u_{10}, u_{11}, c(\cdot), p_1)$ and $\pi^*(u_{00}, u_{01}, u_{10}, u_{11}, c(\cdot), p_1)$, since the boundaries p_L and p_U of the sampling region (p_L, p_U) are themselves functions of payoffs and cost; see (3.8) and (3.9). Moreover, the question of how size and power are related, can now be answered explicitly through their common dependence on u's, c, and p_1.

To illustrate these consequences, we construct the size and power functions of the expected-net-gain-maximizing procedure for deciding between

$$H_0: \omega = \omega_0, \text{ and } H_1: \omega = \omega_1; \quad \omega_1 > \omega_0,$$

where ω is the mean of a normal variate with unit variance. For ease of exposition, suppose $u_{00} = u_{11} = u$, and $u_{01} = u_{10} = 0$. From (3.10) and (3.11),

$$\alpha^*(u,c(\cdot),p_1) = \begin{cases} 1 & ; p_1 \in [0,p_L] \\ 1 - \Phi\left(\dfrac{\log(p_1/(1-p_1))}{\sqrt{n_1^*}(\omega_1-\omega_0)} + \dfrac{\sqrt{n_1^*}(\omega_1-\omega_0)}{2}\right) & ; p_1 \in (p_L,p_U) \\ 0 & ; p_1 \in [p_U,1] \end{cases} \tag{4.1}$$

$$\pi^*(u,c(\cdot),p_1) = \begin{cases} 1 & ; p_1 \in [0,p_L] \\ 1 - \Phi\left(\dfrac{\log(p_1/(1-p_1))}{\sqrt{n_1^*}(\omega_1-\omega_0)} - \dfrac{\sqrt{n_1^*}(\omega_1-\omega_0)}{2}\right) & ; p_1 \in (p_L,p_U) \\ 0 & ; p_1 \in [p_U,1] \end{cases} , \tag{4.2}$$

where Φ is the cumulative distribution function of a standard normal variate: $\Phi(x) = \int_{-\infty}^{x}(1/\sqrt{2\pi})e^{-t^2/2}dt$, and from (3.5),

$$n_1^* = \text{argmax}[-c(n_1) + uE_{\underset{\sim}{x}_1}\{\max(p_2,1-p_2)|p_1,n_1\};\ n_1 \geq 1].$$

This reduces to:

$$n_1^* = \text{argmax}[-c(n_1) + u\{p_1\Psi_0(n_1;p_1) + (1-p_1)\Psi_1(n_1;p_1\};\ n_1 \geq 1], \tag{4.3}$$

where for $k = 0,1$,

$$\Psi_k(n_1;p_1) = \int_{-\log(p_1/(1-p_1))}^{\infty} \frac{e^x p_1/(1-p_1)}{e^x p_1/(1-p_1) + 1} g_k(x)dx$$

$$+ \int_{-\infty}^{-\log(p_1/(1-p_1))} \frac{1}{e^x p_1/(1-p_1) + 1} g_k(x)dx \qquad (4.4)$$

and $g_k(x)$ is the density of a normal variate with mean and variance respectively,

$$\nu_k(n_1) = -(\omega_1-\omega_0)n_1\omega_k + n_1(\omega_1^2-\omega_0^2)/2$$

$$\tau^2(n_1) = n_1(\omega_1-\omega_0)^2.$$

Finally p_L, p_U in (4.1), (4.2) are themselves functions of u, c(•), and p_1 through

$$-c(n_1^*) + u\{p_L\Psi_0(n_1^*;p_L) + (1-p_L)\Psi_1(n_1^*;p_L)\} = u(1-p_L)$$
$$-c(n_1^*) + u\{p_U\Psi_0(n_1^*;p_U) + (1-p_U)\Psi_1(n_1^*;p_U)\} = up_U \;. \qquad (4.5)$$

Calculation of α^*, π^*, n_1^*, p_L, and p_U is usually not so straightforward. Nevertheless for given payoffs, cost, and prior, and for finite discrete-valued data, the combinatorial problem can be solved on the computer. For example, suppose we wish to decide between

$$H_0: \omega = \omega_0, \text{ and } H_1: \omega = \omega_1,$$

where ω is the success probability of a binomial variate based on 4 trials; i.e. data are a random sample from a distribution with probability mass function,

$$f(x;\omega) = \binom{4}{x}\omega^x(1-\omega)^{4-x}.$$

Tables 1 and 2 present values for α^*, π^*, n_1^*, p_L, and p_U, for various values of ω's, u's, c's and p_1's. More specifically, for both tables we have chosen $\omega_0 = 0.5$; $\omega_1 = 0.47, 0.44, 0.41$; $u_{01} = u_{10} = 0$; $c(n) = c.n$. Without loss of generality, we measure expected net gain in units of c, and put $c = 1$. The prior p_1 is allowed to vary over the unit interval in increments of 0.01; checks reveal that this leads to errors in α^*, π^*, in only the fourth decimal place. In Table 1, we have chosen $u_{00} = u_{11} = u = 100$, and in Table 2, $u_{00} = u_{11} = u = 75$.

Table 1 here

Table 2 here

The most obvious feature of the tables is that α^*, π^*, and n_1^* vary with u and p_1. Let us look more closely at the optimal decision procedure as a function of ω_1. The tables show that as ω_1 diverges from ω_0, the sampling interval (p_L, p_U) widens, which can be explained as follows. As $|\omega_1 - \omega_0|$ increases, a first datum offers more discriminatory power between ω_0 and ω_1, thereby increasing the probability of a correct choice with an attendant increase in expected payoff. Whenever this increase exceeds the cost of a datum $c = 1$, sampling will occur. The tables also show that initially, n_1^* tends to increase as $|\omega_1 - \omega_0|$ increases, and the explanation follows similarly. If a second datum's incremental payoff is bigger than its cost $c = 1$, it too is collected, with a higher tendency for this to happen as ω_1 initially moves further from ω_0. Eventually, as ω_1 nears zero or one, n_1^* tends to fall back towards one. This is because a first datum's discriminatory power between ω_0 and ω_1 is then so large that the further improvement provided by a second datum is small, and valued at less than its cost $c = 1$. It is important in all these interpretations to keep in mind this trade-off between payoff and cost.

As one would expect, α^* and π^* decrease as p_1 increases. The reason is that p_2 increases with p_1, thereby increasing the probability of choosing ω_0 at $t = 2 = T$.

5. Discussion

The Neyman-Pearson approach to choosing between a simple null hypothesis $H_0: \omega = \omega_0$, and a simple alternative hypothesis $H_1: \omega = \omega_1$, maximizes the probability of: choosing ω_1 given ω_1 is true, for fixed probability α of: choosing ω_1 given ω_0 is true; that is it maximizes power for fixed size α. Neyman and Pearson (1933) show the optimal test to be based on the likelihood ratio. No mention is made of how to choose the sample size, although the postscript is often added that sufficient observations n_1 are taken so that the optimal size α test achieves a prespecified power π. We call the above type of optimality, NP-optimality.

In this article, we have put the testing problem in a decision theoretic context, and asked that the decision procedure maximize expected net gain. Obviously this is equivalent to minimizing Bayes risk, when the problem is formulated in terms of losses. We have shown that this type of optimality, which we call ENG-optimality, also leads to a procedure based on the likelihood ratio, but one which is sequential in nature and depends on the payoffs, sampling cost, and in particular the prior probability that $\omega = \omega_0$. The size, power, and optimal sample size of this expected-net-gain-maximizing procedure can be calculated; two examples, one continuous and one discrete, are given in Section 4. We shall use the examples to illustrate the link between the two types of optimality.

From the ENG-optimality point of view, fixing the size α and choosing the number of observations in the likelihood ratio test to achieve a prespecified power π, restricts the decisionmaker to payoffs, cost structures and priors that satisfy (4.1) and (4.2). For example, from Table 1, a test of H_0: $\omega_0 = 0.5$, versus H_1: $\omega_1 = 0.41$, with size $\alpha = 17.3\%$ and power $\pi = 50.8\%$, leads to a region for the prior p_1 which in particular does not contain the noninformative prior $p_1 = 1/2$. Therefore restricting size and power to prespecified values severely restricts the prior p_1, when payoffs and costs are held fixed.

A decision procedure which is NP-optimal can clearly be ENG-optimal, provided there is a combination of payoffs, cost structure, and prior which can generate the prespecified α and π. Once such a combination is found, a table like those given would yield the number of observations needed to achieve power π; from (4.1) and (4.2) it would simply be n_1^*.

In explaining the notion of size and power to statistical laymen, there are always difficulties encountered when justifying the particular size, say 5%, that is chosen to perform the test. Imagine the corporate statistician explaining to his or her manager that an investment of millions of dollars, or pounds, etc. should be made based on a test of size 5%. Why not 1%? The manager would more directly comprehend payoffs and costs than probabilities of various types of errors. Our approach of ENG-optimality would provide exactly the type of justification needed to invest or not, and compute rather than arbitrarily specify the probabilities of the various types of errors.

References

Cressie, N. and Morgan, P. B. (1985). The VPRT: a sequential testing procedure dominating the SPRT. The Annals of Statistics, submitted.

Lehmann, E. L. (1959). Testing Statistical Hypotheses. New York: Wiley.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society, A 231, 140-185.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. The Annals of Mathematical Statistics, 19, 326-339.

Table 1

Entries show sampling intervals (p_L, p_U), size α^*, power π^* and sample sizes n_1^* of the optimal decision procedure for choosing between $\omega_0 = 0.5$, and $\omega_1 = 0.47, 0.44, 0.41$, for priors $p_1 \in (p_L, p_U)$, payoffs $u_{00} = u_{11} = 100$, $u_{01} = u_{10} = 0$, and cost function $c(n) = n$.

$\omega_1 = 0.47$: $0.47 < p_L < 0.48$, $0.51 < p_U < 0.52$

	p_1			
	0.48	0.49	0.50	0.51
α^*	0.688	0.688	0.688	0.313
π^*	0.731	0.731	0.731	0.359
n_1^*	1	1	2	1

$\omega_1 = 0.44$: $0.42 < p_L < 0.43$, $0.57 < p_U < 0.58$

	p_1							
	0.43	0.45	0.47	0.49	0.51	0.53	0.55	0.57
α^*	0.588	0.588	0.412	0.412	0.412	0.252	0.252	0.252
π^*	0.779	0.779	0.626	0.626	0.626	0.450	0.450	0.450
n_1^*	4	4	6	5	5	6	5	3

$\omega_1 = 0.41$: $0.35 < p_L < 0.36$, $0.64 < p_U < 0.65$

	p_1							
	0.36	0.40	0.44	0.48	0.52	0.56	0.60	0.64
α^*	0.575	0.425	0.425	0.425	0.286	0.286	0.173	0.173
π^*	0.876	0.782	0.782	0.782	0.656	0.656	0.508	0.508
n_1^*	6	8	8	9	8	10	8	6

Table 2

Entries are as for Table 1, except that $u_{00} = u_{11} = 75$.

$\underline{\omega_1 = 0.47}$: $0.48 < p_L < 0.49$, $0.51 < p_U < 0.52$

	p_1		
	0.49	0.50	0.51
α^*	0.688	0.688	0.313
π^*	0.731	0.731	0.359
n_1^*	1	1	1

$\underline{\omega_1 = 0.44}$: $0.44 < p_L < 0.45$, $0.55 < p_U < 0.56$

	p_1					
	0.45	0.47	0.49	0.51	0.53	0.55
α^*	0.613	0.613	0.387	0.387	0.387	0.194
π^*	0.762	0.762	0.555	0.555	0.555	0.330
n_1^*	2	2	3	3	2	2

$\underline{\omega_1 = 0.41}$: $0.38 < p_L < 0.39$, $0.61 < p_L < 0.62$

	p_1								
	0.39	0.42	0.45	0.48	0.51	0.54	0.57	0.59	0.61
α^*	0.598	0.598	0.402	0.402	0.402	0.227	0.227	0.227	0.227
π^*	0.838	0.838	0.687	0.687	0.687	0.494	0.494	0.494	0.494
n_1^*	4	6	6	5	7	6	5	4	3