Journal of the Korean Statistical Society (2000), **29**: *4, pp 435–441*

Correspondence between Error Probabilities and Bayesian Wrong Decision Losses in Flexible Two-stage Plans[†]

Seoung-gon Ko [1]

ABSTRACT

Ko(1998,1999) proposed certain flexible two-stage plans that could be served as one-step interim analysis in on-going clinical trials. The proposed plans are optimal simultaneously in both a Bayes and a Neyman-Pearson sense. The Neyman-Pearson interpretation is that average expected sample size is being minimized, subject just to the two overall error rates α and β, respectively of first and second kind. The Bayes interpretation is that Bayes risk, involving both sampling cost and wrong decision losses, is being minimized. An example of this correspondence are given by using a binomial setting.

Keywords: Bayes, Neyman-Pearson, Lagrangian optimization, Early termination, Flexible second stage

1. INTRODUCTION

Ko(1998, 1999) proposed a certain two-stage plan with second-stage sample size and critical region dependent on observed first-stage outcome, motivated by the literature of clinical trials - particularly on the issue of the early termination (See for example, O'Brien and Fleming(1979), Armitage(1991)). The development of such plans are based on Neyman-Pearson perspective and however, have also a Bayesian character, much in the way that the sequential probability ratio test simultaneously is optimal in both a Neyman-Pearson and a Bayesian sense(Lehmann(1959)).

The main concerns in this research are to explore this dual Bayes and Neyman-Pearson interpretation of flexible plans with the one-sided one-sample case with

[†]The present research has been conducted in part by the Grant of Kyungwon University in 1999.

[1]Department of Applied Statistics, Kyungwon University, Sungnam 461-701, Korea.

fixed first-stage sample size $m_1 > 0$ and the simple hypotheses $H_0 : p = p_0$ and $H_1 : p = p_1$.

The Bayesian interpretation of flexible two-stage plan was actually almost forced upon us, when we came to realize that what appeared to be a rather ambitious Neyman-Pearson optimization was in fact quite straightforward when see through Bayesian eyes as discussed in Section 2. Section 3 gives the Kuhn-Tucker Lagrangian formulation underlying the optimization and provides an application to the settings in section 2. An example and concluding remarks devote to Section 4.

2. TWO INTERPRETATIONS

Consider the family S of two-stage plans δ with second-stage sample size and critical region dependent on first-stage outcome (say, number of positives) d; here zero second-stage sample size entails not a critical region but rather a terminal decision that may be either "Accept H_0" or "Accept H_1." Our goal, in Neyman-Pearson terms, has been to find a member δ^* of S minimizing average expected sample size $f(\delta)$, where average is taken over H_0 and H_1, subject to given upper bounds on the error rates of the first and second kind. In other words,

$$P_\delta\{Accept\ H_1 \mid H_0\} \leq \alpha \text{ and } P_\delta\{Accept\ H_0 \mid H_1\} \leq \beta.$$

It is natural to set up the problem of finding δ^* in essentially Kuhn-Tucker Lagrangian terms(Spasito(1989)). Suffice it to say here that δ^* will be optimal if there exist positive Lagrange multipliers λ_0 and λ_1 such that

$$g_0(\delta^*) = \alpha,\ g_1(\delta^*) = \beta \text{ and } L(\delta^*) \leq L(\delta), \text{ for all } \delta \in S \tag{1}$$

where $g_0(\delta) = P_\delta\{Accept\ H_1 |\ H_0\ \}$, $g_1(\delta) = P_\delta\{Accept\ H_0|H_1\}$, and $L(\delta) = f(\delta)+\lambda_0 g_0(\delta)+\lambda_1 g_1(\delta)$. Further, the problem of minimizing $L(\delta)$ is separable into autonomous problems, one for each possible first-stage outcome d, each addressing the question of what second-stage sampling plan is best, given that d.

The Bayesian aspect now is seen to enter, when it is realized that the autonomous second-stage problem corresponding to a given first-stage outcome d is in fact a single-sampling Bayes problem, with sampling cost $1/2$, wrong-decision losses λ_0 and λ_1, Bayesian weights on H_0 and H_1. Note especially that λ_0 and λ_1, originally introduced as Lagrange multipliers for the Neyman-Pearson formulation, take on the Bayesian role of economic wrong-decision losses. That is reminiscent of the shadow-price interpretation of dual variables in linear programming(Spasito(1989)); it is also reminiscent of the wrong-decision "weights" constructed in Lehman(1959) for verifying the Bayesian optimality of the sequential probability ratio test.

One further point should be made with regard to computation: The Lagrangian formulation in (1) presupposes that α and β to be specified first, followed by finding appropriate λ_0 and λ_1, along with δ^*. It can be well portrayed by the notation :

$$\lambda_0(\alpha, \beta),\ \lambda_1(\alpha, \beta),\ \text{and}\ \delta(\alpha, \beta).$$

But the reverse computation actually is natural one indicated by the notation :

$$\delta(\lambda_0, \lambda_1),\ \alpha(\lambda_0, \lambda_1)\ \text{and}\ \beta(\lambda_0, \lambda_1).$$

Here one starts out by fixing λ_0 and λ_1, followed by solving for all the second-stage Bayes plans, with wrong-decision losses λ_0 and λ_1, corresponding to all possible first-stage outcomes d, followed in turn by computing the error rates of first and second kind, call them $\alpha(\lambda_0, \lambda_1)$ and $\beta(\lambda_0, \lambda_1)$, of the over-all procedure, call it $\delta(\lambda_0, \lambda_1)$ made up of all of these second-stage Bayes plans. The plan $\delta(\lambda_0, \lambda_1)$ is then the optimal plan δ^* corresponding to the error rate restrictions $g_0(\delta) \leq \alpha(\lambda_0, \lambda_1)$ and $g_1(\delta) \leq \beta(\lambda_0, \lambda_1)$. To meet specified (α, β) objectives, one must iterate the reverse computation till a pair (λ_0, λ_1) are satisfactorily close to α and β.

The reader will recall that we have assumed fixed first-stage sample size $m_1 > 0$. Lacking such a restriction, an optimal procedure must be implemented for successive values of m_2 and the optimum m_1 identified, together with its accompanying δ^*.

3. FORMULATION AND OPTIMIZATION

In this section, we demonstrate the simplest case to apply our Kuhn-Tucker arguments in Section 2 and to deal with the two simple hypotheses (2) concerning a Bernoulli population.

$$H_0\ :\ p\ =\ p_0\ \ vs.\ \ H_1\ :\ p\ =\ p_1(> p_0) \tag{2}$$

These two hypotheses are to be tested against each other according to a two-stage procedure δ. The first stage of δ consists of drawing a sample of size m_1

from the population. While m_1 is in general specified by δ, the chief function of δ, on which we focus in this section, is to specify how the number d, $0 \le d \le m_1$, of first-stage "positive" is to determine the second stage of sampling. In particular, δ is to specify the values of d calling for no additional sampling ($m_2 = 0$), and the values of d, for which m_2 is to be positive. In the former case, δ specifies in addition whether H_0 or H_1 is to be accepted without further sampling. In the latter case, δ specifies in addition the (positive) value of m_2 and the critical region (second-stage numbers D of "positives" leading to acceptance of H_1), to be used after the second-stage sample of size m_2 has been drawn. The dependence of m_2 on d under δ is indicated by the notation $m_{2,\delta}(d)$ or simply $m_\delta(d)$.

From among all possible decision rules δ we propose choosing that rule δ^* which minimizes the average $f(\delta)$ of expected second-stage sample sizes for H_0 and H_1, subject to the restriction that the error rates of first and second kind respectively not exceed predetermined values α and β. When $b(\cdot; m_1, p_0)$ denotes binomial probability, $f(\delta)$ is defined as follows ;

$$2f(\delta) \;=\; \sum_{d=0}^{m_1} m_\delta(d)[b(d;m_1,p_0) \;+\; b(d;m_1,p_1)]$$

Let $R_\delta(d)$ denote the (possibly degenerate) critical D-region to be used under δ, along with $m_\delta(d)$, as a result of observing d, and let $R^c_\delta(d)$ be the complement of $R_\delta(d)$. Then the above restriction is expressed as

$$g_0(\delta) \;=\; \sum_{d=0}^{m_1}[\sum_{D\in R(d)} b(D;m_\delta(d),p_0)]b(d;m_1,p_0) \le \alpha \tag{3}$$

and

$$g_1(\delta) \;=\; \sum_{d=0}^{m_1}[\sum_{D\in R^c_\delta(d)} b(D;m_\delta(d),p_1)]b(d;m_1,p_1) \le \beta \tag{4}$$

Where, in the degenerate case, $m_\delta(d) = 0$, the two square-bracketed terms in (3) and (4) are understood to respectively equal 0 and 1 when d is such that δ calls for immediate acceptance of H_0, and understood to respectively equal 1 and 0 when d is such that δ calls for immediate acceptance of H_1.

Finding δ^* hinges on minimizing the Lagrangian $L(\delta)$ with respect to $\delta \in S$. To this end define $\pi_0(d) = b(d;m_1,p_0)/w(d)$ and $\pi_1(d) = b(d;m_1,p_1)/w(d)$, with $w(d) = b(d;m_1,p_0) + b(d;m_1,p_1)$, and appropriate $\rho_{0,\delta}(d)$ and $\rho_{1,\delta}(d)$ which are depending on δ and d through $(m_\delta(d), R_\delta(d))$. Then write

$$L(\delta) \;=\; \sum_{d=0}^{m_1} w(d)[\pi_0(d)\rho_{0,\delta}(d) + \pi_1(d)\rho_{1,\delta}(d)]$$

The weights $w(d)$ do not depend on δ. And so, as anticipated in section 2, the problem of minimizing $L(\delta)$ with respect to the $(m_1 + 1)$ pairs $(m_\delta(d), R_\delta(d))$ is in fact separable into $(m_1 + 1)$ separate minimization problems, one for each possible value of d, calling respectively for the minimization of the form (5) with respect to (m, R_m), where R_m is a critical region containing none, some or all of the integers between 0 and m.

$$\pi_0[\frac{1}{2}m + \lambda_0 \sum_{D \in R_m} b(D; m, p_0)] + \pi_1[\frac{1}{2}m + \lambda_0 \sum_{D \in R_m^c} b(D; m, p_1)] \tag{5}$$

Also, the task of minimizing (5) with respect to (m, R_m) is in fact the task of finding a single-sample binomial Bayes plan, with weights π_0 and π_1 on H_0 and H_1, which depend on first-stage outcome d, sampling cost 0.5, and wrong decision losses λ_0 and λ_1. These single-sampling Bayes computations, to be carried out for each d are best accomplished by first finding an optimum R_m for fixed m followed by minimizing (5) with respect to m. Finding an optimum R_m for fixed m calls for minimizing

$$\lambda_0 \pi_0 \sum_{D \in R_m} b(D; m, p_0) + \lambda_1 \pi_1 \sum_{D \in R_m^c} b(D; m, p_1)$$

with respect to R_m , yielding an interval of values of D satisfying

$$\lambda_0 \pi_0 \sum_{D \in R_m} b(D; m, p_0) \leq \lambda_1 \pi_1 \sum_{D \in R_m^c} b(D; m, p_1).$$

4. AN EXAMPLE AND CONCLUDING REMARKS

We have looked in detail at an example of the one-sided one-sample case treated in Section 3. For this example with $m_1 = 20$, $(\pi_0, \pi_1) = (\frac{1}{2}, \frac{1}{2})$ and $(p_0, p_1) = (0.15, 0.30)$, we observe $(\lambda_0, \lambda_1) = (1110, 456)$ and best $(\alpha, \beta) = (0.04949, 0.19820)$ when we claim target $(\alpha, \beta) = (0.05, 0.20)$. It is expected because we deal with exact binomial probabilities instead of approximated continuous measure. The resulting plan δ^* is in [Table 1]. Observed (α, β) imply wrong-decision losses (λ_0, λ_1) and given wrong-decisions losses (λ_0, λ_1) imply error rates (α, β). In actuality, then, one need neither a Bayesian nor a Neyman-Pearson be. One should perhaps just look jointly at all four numbers $(\alpha, \beta, \lambda_0, \lambda_1)$, and see if, jointly, they make sense in whatever context is at hand. Essential coincidence of operating characteristic functions between ours and ordinary two-stage plans are observed as expected(Colton and McPherson(1976), Ko(1998, 1999)). Regarding ASN(Average Sampling Number) characteristics, The proposed plan is better uniformly in all range of p, achieving a 8% relative improvement in average expected second-stage sample size at $\bar{p} = (p_0 + p_1)/2$ as in the prior work(Ko(1998)).

There are several articles discussing plans of testing two binomial proportions using approximations rather than exact binomial probabilities(Casagrande and Pike(1978), Feigl(1978)) and Aleong and Bartlett(1979) reconstruct the work of Feigl using Casagrande and Pike's idea. Our work could serve as a benchmark of the plan derived by an approximation and it is fully comparable when we apply their approximation ideas to our plans.

Table 1: The plan δ^* with $p_0 = 0.15$ and $p_1 = 0.30$

d	$M_2(d)$	*Decision*
0	0	A_0^*
1	0	A_0
2	0	A_0
3	0	A_0
4	43	$[11, 43]^{**}$
5	53	$[12, 53]$
6	44	$[9, 44]$
7	0	$A_1{}^{***}$
.	.	.
.	.	.
.	.	.
18	0	A_1
19	0	A_1
20	0	A_1

$*$ Accept p_0

$**$ Critical region for second sample $M_2(d)$

$***$ Accept p_1

References

Aleong J. and Bartlett, D.E., "Improved Graphs for Calculating Sample Sizes when Comparing Two Independent Binomial Distributions," Biometrics, 35, 875-881, 1979

Armitage, P., "Interim analysis in clinical trials," Statistics in Medicine, 10, 925-937, 1991

Casagrande, J.T. and Pike, M.C., "An Improved Approximate Formula for Calculating Sample Sizes for Comparing Two Binomial Distributions," Biometrics, 34, 483-486, 1978

Colton, T. and McPherson, K., "Two-stage plans compared with fixed-sample size and Wald SPRT plans," Journal of the American Statistical Association, 71, 80-96, 1976

Felgl, P, "A Graphical Aid for Determining Sample Size when Comparing two Independent Proportions," Biometrics, 34, 111-122, 1978

Ko, S.G., "Hybrid Group-sequential Conditional-Bayes Approaches to the Double Sampling Plans," The Korean Communications in Statistics, Vol. 5, No. 1, 107-120, 1998

Ko, S.G., "Adjusting Practical aims in Optimal Extended Double Sampling Plans,"The Korean Communications in Statistics, Vol. 6, No. 1, 143-149, 1999

Lehmann, E. L., Testing Statistical Hypotheses, 1st edition, New York, John Wiley and Sons, 1959

O'Brien, P.C. and Fleming, T. R. (1979). "A multiple testing procedure for clinical trials", Biometrics, Vol 35, 549-556.

Sposito, V.A., Linear Programming with Statistical Applications, Iowa State Univ. Press, 1989