MULTI-STAGE TESTS OF HYPOTHESES*

Gary Lorden

California Institute of Technology

A method is given for constructing multi-stage tests that have prescribed error probabilities. The characteristics of the method and its computational requirements are discussed. Numerical examples are given for the case of testing the arrival rate of a Poisson process.

1. Introduction.

Problems of constructing efficient multi-stage tests of hypotheses were studied in Lorden (1983), where it was shown that certain three-stage tests modeled after fully sequential tests are in a typical Koopman-Darmois testing problem asymptotically optimal in comparison with sequential tests. (Two-stage tests are not asymptotically optimal except in degenerate cases.) A major drawback of the tests in that paper is the lack of a good approximation to their error probabilities. The multi-stage tests of the current investigation (whose asymptotic optimality will be established in a later paper) have the advantage of attaining prescribed error probabilities. The price of this feature is additional computational work required to carry out the tests, the amount of work depending on the type of testing problem and the distributions involved.

The multi-stage tests in Lorden (1983) were constructed to emulate the sequential likelihood ratio tests of G. Schwarz (1962), whereas the present approach is, fittingly, modeled after the mixed likelihood ratio tests of

* Research supported by the National Science Foundation under Grant DMS-8301273.

AMS 1980 subject classification. Primary 62F05; Secondary 62L10.

Key words and phrases. Multi-stage hypothesis tests, asymptotic efficiency.

Robbins (1970), with precursors in the work of Ville (1939) and Wald (1947). Although the method of construction used here applies to the open-ended testing problems considered by Robbins, it is more revealing to consider another special case, the problem of testing between two simple hypotheses, $\theta = \theta_0$ and $\theta = \theta_1$, about the parameter θ of a dominated family of distributions. The natural Robbins-type mixture rules for this problem are based upon a ("prior") probability measure, G, on Θ, the parameter space, and the induced mixture of the probabilities $P_\theta(.)$ on the space of sample sequences, which will be denoted by $Q(.)$. Using the notation $(dQ/dP_i)_n$ to denote the G-mixed likelihood ratio of the P_θ's to P_{θ_i} after n observations, the fully sequential tests stop at N = the smallest $n \geq 1$ such that

$$(1) \qquad (dQ/dP_i)_n \geq A_i \text{ for } i=0 \text{ or } 1,$$

with appropriate rejections, the A_i's being chosen in advance.

2. Definition and properties of the tests.

Let $n(\alpha_0,\alpha_1)$ denote the smallest fixed sample size for which error probabilities α_0,α_1 or smaller can be achieved. Assume that the desired α_0,α_1 are given and that $n(\alpha_0,\alpha_1)$ is greater than one.

Determination of the first stage: Choose the design parameter $0 < \gamma < 1$, whose role will be to fix

$$(2) \qquad Q(\text{stop at first stage}) = \gamma.$$

Find the first stage sample size, m, such that for some critical values C_i, i=0,1, the regions

$$(3) \qquad \text{reject } \theta_i \text{ if } (dQ/dP_i)_m \geq C_i$$

with possible randomization in case of equality, yield

$$(4) \qquad \alpha_i^* := P_i(\text{reject } \theta_i \text{ at first stage}) = \gamma\alpha_i, \quad i = 0,1,$$

and also satisfy (2). Note that it is in general necessary to randomize between two consecutive values of m in order to satisfy (2) and (4) (see the discussion in the next section). It is, however, often unnecessary to do any calculations involving (3), since the same rejection regions may be specifiable in terms of a convenient sufficient statistic.

Construction of the continuation (if required): Determine the (conditional) error probabilities

$$(5) \qquad \tilde{\alpha}_i = \alpha_i(dQ/dP_i)_m, \quad i = 0,1,$$

for the continuation, select a new design parameter, $\tilde{\gamma}$, and apply the same prescription given for the first stage, with the role of Q now played by the mixture $\tilde{Q}$ induced by the posterior, $\tilde{G}$. The design parameters, γ, $\tilde{\gamma}$, etc. are conveniently chosen before starting the test, e.g. for a three-stage test one could specify that $\gamma = .5, .4, 1$, say, for the respective stages (the intended last stage necessarily having $\gamma = 1$).

From (3) it is clear that whenever the test continues

$$(dQ/dP_i)_m < E_i[(dQ/dP_i)_m | \text{reject } \theta_i \text{ at first stage}]$$

for $i = 0,1$ and, hence,

$$\begin{aligned}
\sum_i \tilde{\alpha}_i &< \sum_i \alpha_i E_i[(dQ/dP_i)_m \mid \text{reject } \theta_i \text{ at first stage}] \\
&= \sum_i \alpha_i Q(\text{reject } \theta_i \text{ at first stage})/\alpha_i^* \\
&= \sum_i Q(\text{reject } \theta_i \text{ at first stage})/\gamma \\
&= 1.
\end{aligned}$$

The same is clearly true at all later stages.

It is easy to verify that the desired error probabilities are achieved:

$$(8) \qquad P_i(\text{reject } \Theta_i) = \alpha_i^* + E_i\tilde{\alpha}_i 1\{\text{continue}\}$$

$$= \gamma\alpha_i + E_i\alpha_i(dQ/dP_i)_m 1\{\text{continue}\}$$

$$= \alpha_i(\gamma + Q(\text{continue})) = \alpha_i.$$

As an illustration of the type of computation required in simple cases, consider the problem of testing whether the drift per unit time, Θ, of a (unit variance) Wiener process $\{X(t)\}$ equals -1 or $+1$. Then it is easy to see that the first-stage error probabilities α_i^* in (4) are attained after sampling for (fixed) time t by the rule

$$\text{reject } \Theta_0 = -1 \text{ if } X(t) \geq z_0^* t^{1/2} - t,$$

$$\text{reject } \Theta_1 = 1 \text{ if } X(t) \leq z_1^* t^{1/2} + t,$$

$$\text{continue otherwise,}$$

where z_o^* and z_1^* denote the upper α_0^* quantile and the lower α_1^* quantile, respectively, of the standard normal distribution. The value of the first-stage sample size, t, is determined by "trial and error" to satisfy

$$0 = h(t) = 1 - \gamma - Q(\text{continue after the first stage})$$

$$= 1 - \gamma - \int (\Phi(z_0^* - (\Theta + 1)t^{1/2}) - \Phi(z_1^* - (\Theta - 1)t^{1/2}))G(d\Theta),$$

where Φ denotes the standard normal distribution function. The calculation of values $h(t)$ is easy of G is normal or is supported on a finite set. (For the

discrete time testing of a normal mean, only integer t are used and either one has to settle for an approximate zero of h(t) or else randomize the choice of t.) If the first stage fails to stop, then a straightforward calculation of the conditional error probabilities, $\tilde{\alpha}_i$, can be made using the G-mixture of likelihood ratios; the same calculations (with different α_i^*) are used to construct the second stage and later stages.

3. A Poisson example.

Suppose that one observes a Poisson process $\{N(t),\ 0 \leq t < \infty\}$ with arrival rate θ and it is desired to test $\theta = \theta_0$ vs. $\theta = \theta_1 > \theta_0$. (This problem permits simple calculation of expected sample sizes.) Let the "prior" G have (natural conjugate) gamma density

$$\frac{\beta^\nu \theta^{\nu-1}}{\Gamma(\nu)} e^{-\beta\theta},\ \theta > 0,$$

where $\nu, \beta > 0$. Then under the G-mixture, Q, the number of arrivals in time t has negative binomial density

$$f(n;\nu,p) = \binom{\nu+n-1}{n} p^\nu (1-p)^n,\ n=0,1,\ldots, \tag{9}$$

where $p = \beta/(\beta+t)$. For randomization purposes it is helpful to regard the probability mass $f(n;\nu,p)$ as "smeared evenly" over $[n,n+1)$, yielding the distribution functions

$$F(x;\nu,p) = \sum_{n=0}^{[x]} f(n;\nu,p) - (x+1-[x])f([x];\nu,p),\ x \geq 0. \tag{10}$$

Similarly, let $h(n;\lambda)$ denote the Poisson probability $\exp(-\lambda)\lambda^n/n!$ and define the "smeared" distribution function

$$H(x;\lambda) = \sum_{n=1}^{[x]} h(n;\lambda) - ([x]+1-x)h([x];\lambda),\ x \geq 0. \tag{11}$$

For typical choices of the prior parameters ν and β, the construction of the first stage proceeds as follows. The form of the rejection rules (3) for first stage sampling time τ becomes

(3') reject θ_0 if $N(\tau) \geq b$, reject θ_1 if $N(\tau) \leq a$,

where it is convenient (for randomization) to allow a and b to be arbitrary non-negative numbers under the following interpretation: if $N(\tau) = [a]$ then reject θ_1 with probability $a - [a]$, whereas if $N(\tau) = [b]$ then reject θ_0 with probability $[b] + 1 - b$. To determine a and b such that (4) is satisfied, express it as

(4') $$H(a;\theta_1\tau) = \gamma\alpha_1 \text{ and } 1-H(b;\theta_0\tau) = \gamma\alpha_0,$$

which determines a and b as "Poisson quantiles" for given τ. To satisfy relation (2), note that

$$Q(\text{stop at first stage}) = 1 - [F(b;\nu,\beta/(\beta+\tau)) - F(a;\nu,\beta/(\beta+\tau))].$$

This last increases in τ, so that finding the τ that yields the value γ required by (2) is straightforward numerically. This first-stage sampling time, τ, and its associated a and b determine the first stage of the test. Continuation, if called for, is based upon the same constructions, starting with

$$\tilde{\alpha}_i = \alpha_i f(N(\tau);\nu,\beta/(\beta+\tau))/h(N(\tau);\ \theta_i\tau), \quad i=0,1,$$

and the posterior $\tilde{G}$ and its $\tilde{Q}$, which are determined by $\tilde{\beta} = \beta + \tau$ and $\tilde{\nu} = \nu + N(\tau)$.

Note that in the more common discrete time situations there are, in effect, only integer values of τ available, so that Q (stop at first stage) generally "crosses over" γ at consecutive values τ and $\tau + 1$. Randomization of the choice between τ and $\tau + 1$ for the first stage sample size yields the

corresponding convex combination of the Q(stop)'s for τ and $\tau + 1$, which is easily chosen to yield γ.

A series of computer calculations was made to compare the expected sample sizes of tests of the type proposed with those of fixed sample size tests and 2-SPRT's. The latter sequential tests have triangular continuation regions and were shown in Lorden (1976) to nearly minimize the expected sample size at the specified θ chosen in their construction. The results shown below are typical of those obtained over a range of error probabilities in the problem of testing $\theta = 1$ vs. $\theta = 2$. The value $\theta = 1.5$ was used to construct the 2-SPRT and was included with $\theta = 1$ and 2 in the evaluation of sample sizes. The parameters chosen for the gamma prior were $\nu = 60$, $\beta = 40$, yielding a mean of 1.5 and standard deviation of .19. The values of the design parameter γ in the multiple stages were chosen (without an exhaustive search) to make the expected sample size small at $\theta = 1.5$.

Expected Sample Sizes
(α_0=4.93%, α_1=4.83%)

	θ=1	θ=1.5	θ=2
fixed sample	16.1	16.1	16.1
2-stage (γ=.63,1)	12.1	13.9	11.7
3-stage (γ=.45,.6,1)	11.1	13.2	10.4
2-SPRT	9.7	11.2	7.5

4. General testing problems.

Suppose that k terminal decisions are possible and let θ_i, i=1,...,k, be the subset of the parameter space where the i^{th} decision incurs positive regret. In addition to G, one needs to specify for each i a mixing measure G_i supported on θ_i. Letting $P_i(.)$ denote the induced measure, the error probability constraints become

$$P_i(\text{decision } i) = \alpha_i, \quad i=1,\ldots,k. \tag{12}$$

(In the θ_0 vs. θ_1 problem, let the i^{th} decision be "reject θ_i" and let G_i concentrate at θ_i.) Then (1) defines sequential tests of the asymptotically optimal type studied by Kiefer and Sacks (1963). The definition and properties

of the multi-stage tests given in Section 2 apply with the changes that "reject θ_i" becomes "decision i" everywhere and after (5) is used in the construction of the continuation, the role of P_i is played by $\tilde{P}_i$, induced by the "posterior", $\tilde{G}_i$, associated with G_i.

As an example of this formulation, consider the problem of testing $k \geq 3$ simple hypotheses (with G possibly supported on an interval). Identify "decision i" as "choose θ_i" (which is different from the $k = 2$ description above, where the decision labels are the reverse), so that G_i assigns weights to the θ_j's with $j \neq i$ and α_i equals the corresponding weighted average of probabilities of choosing θ_i.

As another example, consider the problem of testing $\theta = \theta_0$ vs. $\theta \neq \theta_0$. Let G_0 be concentrated at θ_0 and G_1 in $\theta - \theta_0$ (perhaps at two points), with the error constraints controlling P_{θ_0} (reject θ_0) and the G_1-mixture of P_θ (accept θ_0). In typical Koopman-Darmois cases, the computational work required to satisfy (2) and (4) in the construction of each stage will be greater than in the θ_0 vs. θ_1 problem, but still reasonable.

REFERENCES

Kiefer, J. and Sacks, J. (1963). Asymptotically optimum sequential inference and design. Ann. Math. Statist. **34** 705-750.

Lorden, G. (1976). 2-SPRT's and the modified Kiefer-Weiss problem of minimizing an expected sample size. Ann. Statist. **4** 281-291.

Lorden, G. (1983). Asymptotic efficiency of three-stage hypothesis tests. Ann. Statist. **11** 129-140.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. Ann. Math. Statist. **41** 1397-1409.

Schwarz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. Ann. Math. Statist. **33** 224-236.

Ville, J. (1939). Étude Critique de la Notion de Collectif. Gauthier-Villars, Paris.

Wald, A. (1947). Sequential Analysis. Wiley, New York.