The Annals of Statistics
1981, Vol. 9, No. 2, 461–464

APPLYING ASYMPTOTIC SHAPES TO NONEXPONENTIAL FAMILIES

By Gideon Schwarz

Hebrew University, Jerusalem

Fortus' generalization of asymptotic shapes of optimal testing regions for composite hypotheses does away with the restriction to exponential families originally imposed by us. Here we survey his work critically, and suggest some improvements that may be crucial for its practical applicability to parametric problems, and point out its shortcomings for nonparametric ones.

1. Introduction: asymptotic shapes and Fortus' generalization. Asymptotic shapes were introduced in an earlier paper (1962). They arise as follows. First, when a statistical hypothesis H_0 is to be tested against an alternative H_1, on the basis of sequentially sampled independent observations that cost c units each, the optimal procedure is related to the posterior stopping risk R. When the latter reaches a value less than c, the optimal (Bayes) procedure will obviously call for stopping; for "separated" hypotheses, we have shown also (for some c_0 and K) that as long as R exceeds $K\,c\,\log(1/c)$ where $c < c_0$, the optimal procedure leads to taking another observation. For these facts, that can be conveniently expressed as inclusions of events

$$\{R < c\} \subset \{\text{optimal procedure stops}\} \subset \left\{R \le K\,c\,\log\frac{1}{c}\right\},$$

no further assumptions are required.

For the second step, the distributions of the observations were assumed to form a (k-dimensional) exponential family. For this case, the three events forming the chain of inclusions above can be interpreted as sets in the $(k+1)$-dimensional space of $S(X_1) + \cdots + S(X_n)$, the (k-dimensional) sufficient statistic of the first n observations, with n itself forming the k+first coordinate. It was then shown that, as $c \to 0$, the two sets at the ends of the chain grow at the rate of $\log(1/c)$, and if this growth is counteracted by shrinking them at that rate, both tend to the same limit-set, and hence, the same holds for the optimal stopping set sandwiched between them, if it too is rescaled by shrinking it $\log(1/c)$-fold. The limit-set is the *asymptotic shape*, and blowing it up back to $\log(1/c)$ times its size yields an approximation to the optimal stopping set. In terms of the generalized likelihood ratio statistic Λ_ι for testing H_ι against its complement, the approximate stopping region is the set where at least one of Λ_0 and Λ_1 exceeds $1/c$.

Recently Fortus (1979) attempted to do away with the restriction to exponential families. As those are characterized by the existence of a vector valued statistic that is sufficient when summed over the observations, Fortus chose a function valued statistic to play a similar role: the log likelihood function. In the linear (∞-dimensional) space of these functions, with one dimension added for n, stopping regions and regions of constant posterior risk are well defined. The concept of shrinking (by $\log(1/c)$) is meaningful here as well, and so asymptotic shapes are obtained, and the approximate procedure that results from replacing the actual shape by the asymptotic shape is defined by Fortus just as in the exponential-family case, and can be expressed in terms of Λ_0 and Λ_1 here as well.

An important improvement added by Fortus to his generalization, is the proof of local uniformity of the convergence of the scaled region to its asymptotic shape.

Received October 1979.
AMS 1970 *subject classification.* 62L15, 62F05.
Key words and phrases. Asymptotic shapes, optimal sequential tests.

2. Interpreting the convergence and its uniformity. In the exponential case, when the stopping risk R is regarded as a function of $\Sigma = S(X_1) + \cdots + S(X_n)$ and n, the domain of this function consists properly of those pairs (Σ, n) which are attained by some possible sequence $X_1, \cdots, X_n$. It is convenient to extend R to all pairs where its formula is meaningful. We (1962) mentioned one part of this extension (the inclusion of noninteger n values) but failed to mention that in some cases, such as for integer-valued S, we assumed R to be defined as if also S were real valued. Fortus (1979) proceeds likewise.

Only with the domain thus extended is the geometric description of the various regions and shapes valid, and this must be kept in mind when one attempts to evaluate one characteristic feature of the asymptotic shapes method: as n tends to infinity, the mean sufficient statistic T_n (equal Σ/n in the exponential case and the log likelihood divided by n in Fortus' case) is held fixed.

For the convergence of the scaled regions to the asymptotic shape, the fixing of T_n is merely a technical device, made appropriate by the fact that the regions grow in all directions at the same asymptotic rate of $\log(1/c)$ when c tends to zero. However, when the asymptotic shape is to be used in a real problem, where c is small but positive, T_n will never be fixed as n increases, and the justification of using the approximate procedure depends on two further results. One is the local uniformity in T_n of the convergence. This result is new in Fortus (1979) even for the exponential case. But, to utilize the local uniformity in T_n for an evaluation

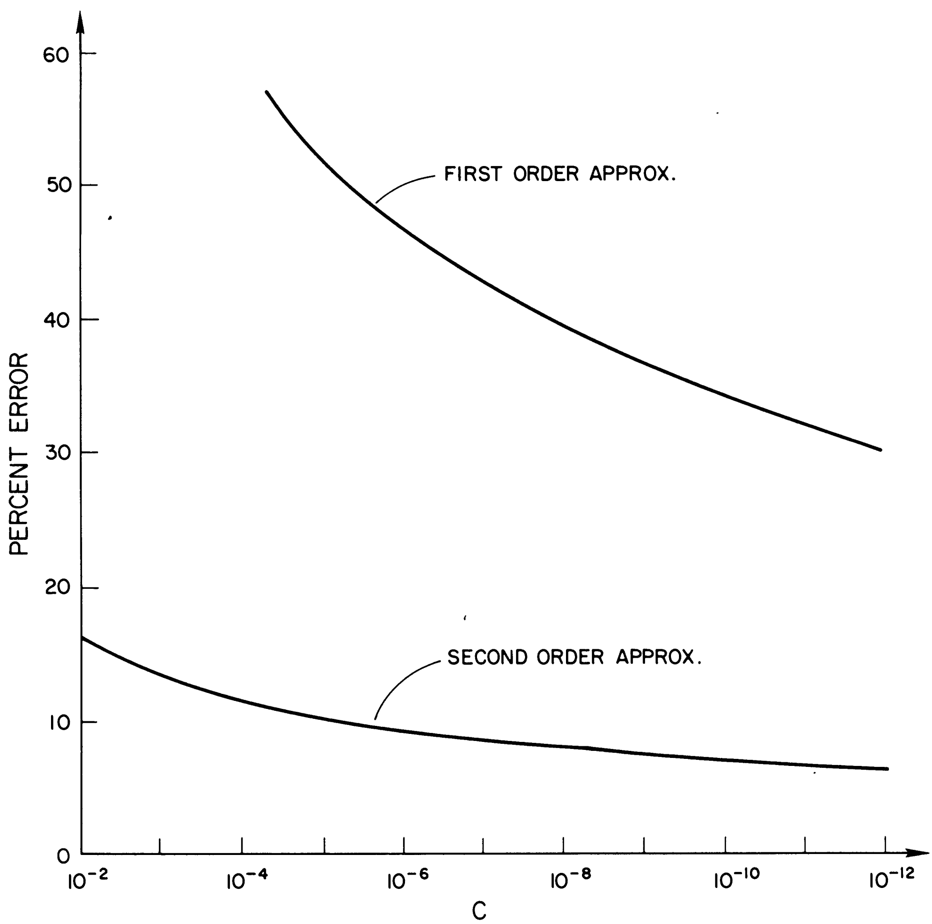

FIG. 1

of the asymptotic procedure, T_n must be shown to remain in a set for which the uniformity is valid, as sampling proceeds.

For interior parameter points of an exponential family, the required behavior of T_n is guaranteed by the law of large numbers: there T_n is the mean of n independent identically distributed vectors $S(X_i)$, with finite moments of all orders; consequently it converges almost surely to its expectation. In Fortus' case, T_n is function valued, and even pointwise, $T_n(\theta)$ may not converge. For $E(T_n(\theta))$ is the Kullback-Leibler distance between the uniform distribution and the distribution under θ, and this distance may be infinite. In such a case T_n will not stay in a bounded set, and the uniformity will not apply.

We therefore add one condition to Fortus' assumptions: for $X_1, \ldots, X_n, \ldots$ independent, distributed according to one of the distributions in the parameter space, the mean log likelihood $T_n(\theta|X_1, \ldots, X_n)$ stays almost surely in a set (of functions) that is bounded (in the metric $\sup|\exp f - \exp g|$); or equivalently, the set

$$\{\exp(T_n(\theta|X_1, \ldots, X_n)) - \exp(T_m(\theta|X_1, \ldots, X_m))\}$$

is almost surely bounded.

Whenever this condition holds, Fortus' description of the approximation as "reasonable ... for small c" can be justified by applying his uniformity result. For the practical application, there is still the question how small is "small".

3. The second order correction to the size. At the end of his paper, Fortus quotes Fushimi (1967), who found in numerical examples that for $c = 10^{-8}$ the approximation is still far from reasonable. The limitation imposed thereby on the application is seen to be less severe if one considers that c is the cost of an observation in units of the penalty for a wrong decision, and that $c = 10^{-8}$, e.g., corresponds to sample sizes of the order of magnitude of log 10^8, which is less than 20.

Fushimi proceeds to find a second order correction for the one-dimensional normal case, with linear loss; subsequently we generalized it to other one-dimensional exponential families and other loss functions (1969). In one sense these results are incomplete: the second order corrections for the two regions that flank the optimal region in the chain of inclusions in Section 1 differ from each other asymptotically by log log (1/c), and therefore the optimal region cannot be approximated by this method any closer than log log (1/c). This is also the order of magnitude of the correction term, so not much seems to be gained by including it. Still, using it one can approximate the optimal stopping region with an error term equal to ½ log log (1/c) + O (1), while without it, the error contains higher multiples of the log log term, i.e., at least 3/2 log log (1/c) in the case treated by Fushimi. Since the regions are in (Σ,n)-space, the error mentioned above corresponds to an error proportional to log log (1/c) in sample size, or to c log log (1/c) in cost. In either description, the relative error is asymptotically ½log $\log(1/c)/\log(1/c)$. For the one-dimensional case with losses proportional to the squared distance from the indifference region, the relative error would be five times as large, if the second order correction were ignored (see figure).

For applications, the second order correction is clearly of crucial importance. Since it varies with the dimension of the exponential family, no one form will do for the general case. In fact, since it grows proportionally with the dimension, it appears most necessary, yet least accessible when the dimension becomes S infinite, as it may be under Fortus' assumptions. It can be salvaged, however, if we retain an assumption of finite-dimensionality less stringent than that of an exponential family. In the latter, the dimension of the parameter space is also the linear dimension of the log densities. The second order correction terms generalize under some regularity assumptions to the case where the parameter space is Euclidean k-space, as we now proceed to exemplify by the case $k = 1$.

So, we let θ be real valued, and strengthen Fortus' continuity assumption by requiring the likelihood function to be unimodal and to possess bounded second derivatives, a condition that holds automatically in the exponential case. Also, we assume the hypotheses to be halflines separated by a finite interval (θ_0,θ_1), and the loss function to be bounded, and to behave like $|\theta-\theta_i|^\rho$ just outside the interval. Finally, we assume an a priori density, bounded between positive numbers in every finite interval.

Under these assumptions, the evaluation of the second order correction in Schwarz (1969) goes through, and yields for the size factor by which to blow up the asymptotic shape

$$\log\frac{1}{c} - \left(\rho + 1 \pm \frac{1}{2}\right)\log\log\frac{1}{c}\,.$$

Thus corrected, Fortus' generalization yields an approximation applicable in the case of a parametric family. For nonparametric problems, though formally correct, the approximation cannot be corrected, and without a correction it remains too rough to be of any practical value.

For exponential families, the gap between the constant-risk bounds of the Bayes regions has been eliminated by Lorden (1967, 1977, 1980) who showed that for appropriate M^*, the Bayes procedure does not stop as long as R exceeds M^*c. This determines the correct sign preceding the ½ in the last formula to be a minus, and reduces the relative error to $O((\log c^{-1})^{-1})$. Hopefully this result, that is best possible if full dependence on the prior is avoided, can also be extended beyond exponential families.

REFERENCES

FORTUS, ROBERT (1979). Approximations to Bayesian sequential tests of composite hypotheses. *Ann. Statist.* **7** 579–591.

FUSHIMI, M. (1967). On the rate of convergence of asymptotically optimal Bayes tests. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* **14** 1–7.

LORDEN, GARY (1967). Integrated risk of asymptotically Bayes sequential tests. *Ann. Math. Statist.* **38** 1399–1422.

LORDEN, GARY (1977). Nearly-optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5** 1–21.

LORDEN, GARY (1980). Nearly-optimal sequential tests for exponential families. Unpublished manuscript.

SCHWARZ, GIDEON (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

SCHWARZ, GIDEON (1969). A second order approximation to optimal sampling regions. *Ann. Math. Statist.* **70** 313–315.

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM, ISRAEL