Pattern Recognition, Vol. 23, No. 11, pp. 1275–1279, 1990
Printed in Great Britain

EFFICIENCY OF LOGISTIC-NORMAL STOCHASTIC SUPERVISION

T. KRISHNAN and S. C. NANDY
Indian Statistical Institute, 203, Barrackpore Trunk Road, Calcutta 700 035, India

(*Received 8 August* 1989; *received for publication* 7 *December* 1989)

Abstract—We consider the case of a stochastic supervision scheme in which the logarithm of odds ratio of the supervisor classification has a normal distribution and is independent of the feature vector. We consider the two-group multinormal case with a common covariance matrix and compute the Efron efficiency of this supervision scheme, for various values of parameters of the feature distribution and the supervision distribution. We interpret these efficiency values to draw conclusions on the worth of stochastic supervision.

Discriminant analysis Stochastically classified initial samples Asymptotic relative efficiency Logistic-normal distribution

1. INTRODUCTION

Applications in medical diagnosis and remote sensing have stimulated studies of pattern recognition problems with imperfect supervision. In a series of papers in *Pattern Recognition*[1–4] we developed algorithms for learning a mixture of two p-dimensional normal distributions with a common covariance matrix, under imperfect supervision of two types: (i) deterministic but error-prone and (ii) stochastic; we also computed the efficiency of these schemes relative to perfect supervision. In our studies of stochastic supervision we used the beta distribution as a model for the supervisor's assessment. In a recent article, Titterington[5] proposed a logistic-normal model for the supervisor assessment, namely, normal distribution for logoddsratio of the supervisor assessment, showed its flexibility and similarity with our beta model and developed the EM algorithm for maximum likelihood estimation of parameters. In this article, we study the Efron efficiency of the Titterington-model-based stochastic supervision scheme.

Let:
Z: Supervisor assessment of the probability of a unit belonging to group 1;
W: $\log[Z/(1-Z)]$;
X: a p-dimensional feature vector.

Then, we assume that the distribution of W is normal with mean $-\eta, \eta$ respectively in the groups 0 and 1, and with a common variance σ^2 in the two groups. The feature vector X has p-dimensional normal distributions with mean vectors μ_0, μ_1 in groups 0 and 1 respectively and with a common dispersion matrix Σ in the two groups. Further, we assume that, in each group, X and Z are independent. The prevalence rates of the two groups are π_0 and π_1 respectively. An initial sample of N units is taken from the mixture

$$\pi_0 f_0(x, z) + \pi_1 f_1(x, z) \tag{1.1}$$

where $f_0(x, z)$ and $f_1(x, z)$ are obtained as described above.

Efron's[6] efficiency is a measure of the information contained in an initial sample unit taken according to a certain scheme compared to a perfectly supervised scheme; another way of looking at Efron efficiency is as the relative sample size required under a certain scheme compared to perfect supervision to achieve the same estimation efficiency of the discriminant function coefficients based on the feature vector X. The mathematical definition of Efron efficiency and its derivation and uses in various cases can be found in many references and hence we avoid going into its details here—the definition and use in studying the efficiency of logistic regression can be found in Efron,[6] use in studying the efficiency of unsupervised schemes in O'Neill,[7] efficiency of error-prone deterministic supervision schemes in Krishnan[2] and stochastic supervision schemes with a beta supervisor in Krishnan and Nandy.[4]

It is easily seen that the model above is an unsupervised scheme based on a $(p+1)$-dimensional feature vector (X, Z) where the vector has normal distributions in each group with a common covariance matrix; this common covariance matrix has a special structure whereby the covariances of Z with the components of X are all zero. This is a special case of the unsupervised scheme of O'Neill's.[7] While a great deal of effort was required to derive formulae for the Efron efficiency of the beta supervisor, under the Titterington logistic-normal supervisor it becomes easy owing to the normality of (Z, X).

The Efron efficiency based on a p-dimensional feature vector is a convex combination of what Efron calls the intercept and slope efficiencies of the discriminant boundary which are also interpretable respectively as the efficiencies when $p = 1$ and $p = \infty$.

Efron efficiency depends on π_1 and the Mahalanobis distance between the two groups. We quote below O'Neill's formulae for efficiency in the case of a p-dimensional vector, denoted Eff_p, with a distance of Δ between the two groups.

Let $\lambda = \log \dfrac{\pi_1}{\pi_0}$.

$$\mathrm{Eff}_p(\pi_1, \Delta) = \frac{q(\pi_1, \Delta)\mathrm{Eff}_1(\pi_1, \Delta) + (p-1)\,\mathrm{Eff}_\infty(\pi_1, \Delta)}{q(\pi_1, \Delta) + (p-1)} \qquad (1.2)$$

where

$$q(\pi_1, \Delta) = \frac{\left(1, -\frac{\lambda}{\Delta}\right)[H - A]^{-1}\left(1, -\frac{\lambda}{\Delta}\right)' \mathrm{Eff}_\infty}{1 + \pi_1\pi_0\Delta^2}; \qquad (1.3)$$

$$\mathrm{Eff}_1(\pi_1, \Delta) = \frac{\left(1, -\frac{\lambda}{\Delta}\right) H^{-1}\left(1, -\frac{\lambda}{\Delta}\right)'}{\left(1, -\frac{\lambda}{\Delta}\right)[H - A]^{-1}\left(1, -\frac{\lambda}{\Delta}\right)'}; \qquad (1.4)$$

$$\mathrm{Eff}_\infty(\pi_1, \Delta) = 1 - a_0(1 + \pi_0\pi_1\Delta^2); \qquad (1.5)$$

where

$$H^{-1} = \begin{bmatrix} (1 + \Delta^2/4 & -(\pi_0 - \pi_1)\Delta/2 \\ -(\pi_0 - \pi_1)\Delta/2 & 1 + 2\pi_0\pi_1\Delta^2 \end{bmatrix}; \qquad (1.6)$$

$$a_i = \int_{-\infty}^{\infty} x^i \frac{\exp[-\Delta^2/8]\phi(x)}{\pi_1 \exp[\Delta x/2] + \pi_0 \exp[-\Delta x/2]}\,\mathrm{d}x; \qquad (1.7)$$

$$A = \begin{pmatrix} a_0 & a_1 \\ a_1 & a_2 \end{pmatrix}. \qquad (1.8)$$

In our case the distance between the two groups on the basis of the $(p + 1)$ dimensions denoted by ∇ is given by

$$\nabla^2 = \Delta^2 + \delta^2, \qquad (1.9)$$

where Δ is the distance between the two groups based on X and δ is the distance based on Z and $\delta^2 = \dfrac{4\eta^2}{\sigma^2}$. Thus the efficiency formulae in our case are obtained by plugging in ∇ as given by equation (1.9) instead of Δ in formulae for Eff as given above.

We denote our stochastic supervision efficiency based on a p-dimensional feature vector by $\mathcal{S}\mathrm{Eff}_p$ although by O'Neill's notation it is Eff_{p+1} being unsupervised efficiency for a $(p + 1)$-dimensional feature vector. Thus

$$\mathcal{S}\mathrm{Eff}_p(\pi_1, \Delta, \delta) = \mathrm{Eff}_{p+1}(\pi_1, \nabla). \qquad (1.10)$$

When $\delta = 0$,

$$\mathcal{S}\mathrm{Eff}_p(\pi_1, \Delta, 0) = \mathrm{Eff}_{p-1}(\pi_1, \Delta); \mathcal{S}\mathrm{Eff}_1(\pi_1, \Delta, 0) = \mathrm{Eff}_2(\pi_1, \Delta). \qquad (1.11)$$

In the following section we present a table and a chart of this $\mathcal{S}$Eff for various parameter values.

2. COMPUTATION OF EFFICIENCY AND INTERPRETATION

We have computed $\mathcal{S}\mathrm{Eff}_1$ and $\mathcal{S}\mathrm{Eff}_\infty$ as given by the formulae derived above for various values of π_1, Δ and δ. Efficiency values for some selected values of the parameters are presented in Table 1. A summary of the efficiencies is presented in Fig. 1.

From Table 1 and Fig. 1, we notice that the results are similar to what we obtained for the beta supervision in our earlier paper,[4] namely that:

1. Efficiency increases with supervision, i.e. with increasing δ;
2. Efficiency increases with Δ, the distance between the two groups;
3. For $\pi_1 = 0.5$, the dimension of the feature vector is immaterial; this also clearly follows from the formulae for efficiency;

Table 1. Asymptotic relative efficiency of normal discrimination with stochastic (logistic-normal) supervision

$\pi_1 = 0.5$

	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
δ	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$
0	0.1008	0.1008	0.3594	0.3594	0.6589	0.6589
2	0.3072	0.3072	0.5467	0.5467	0.7716	0.7716
3	0.5467	0.5467	0.7201	0.7201	0.8652	0.8652
4	0.7716	0.7716	0.8652	0.8652	0.9376	0.9376

$\pi_1 = 0.667$

	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
δ	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$
0	0.0989	0.1211	0.3572	0.3819	0.6561	0.6722
2	0.3051	0.3305	0.5441	0.5642	0.7690	0.7805
3	0.5441	0.5642	0.7174	0.7311	0.8630	0.8702
4	0.7690	0.7805	0.8630	0.8702	0.9362	0.9397

$\pi_1 = 0.9$

	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
δ	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$	$\mathcal{S}\mathrm{Eff}_1$	$\mathcal{S}\mathrm{Eff}_\infty$
0	0.0798	0.1994	0.3120	0.4896	0.6134	0.7481
2	0.2624	0.4384	0.4976	0.6570	0.7334	0.8345
3	0.4976	0.6570	0.6781	0.7957	0.8364	0.9031
4	0.7334	0.8345	0.8364	0.9031	0.9199	0.9548

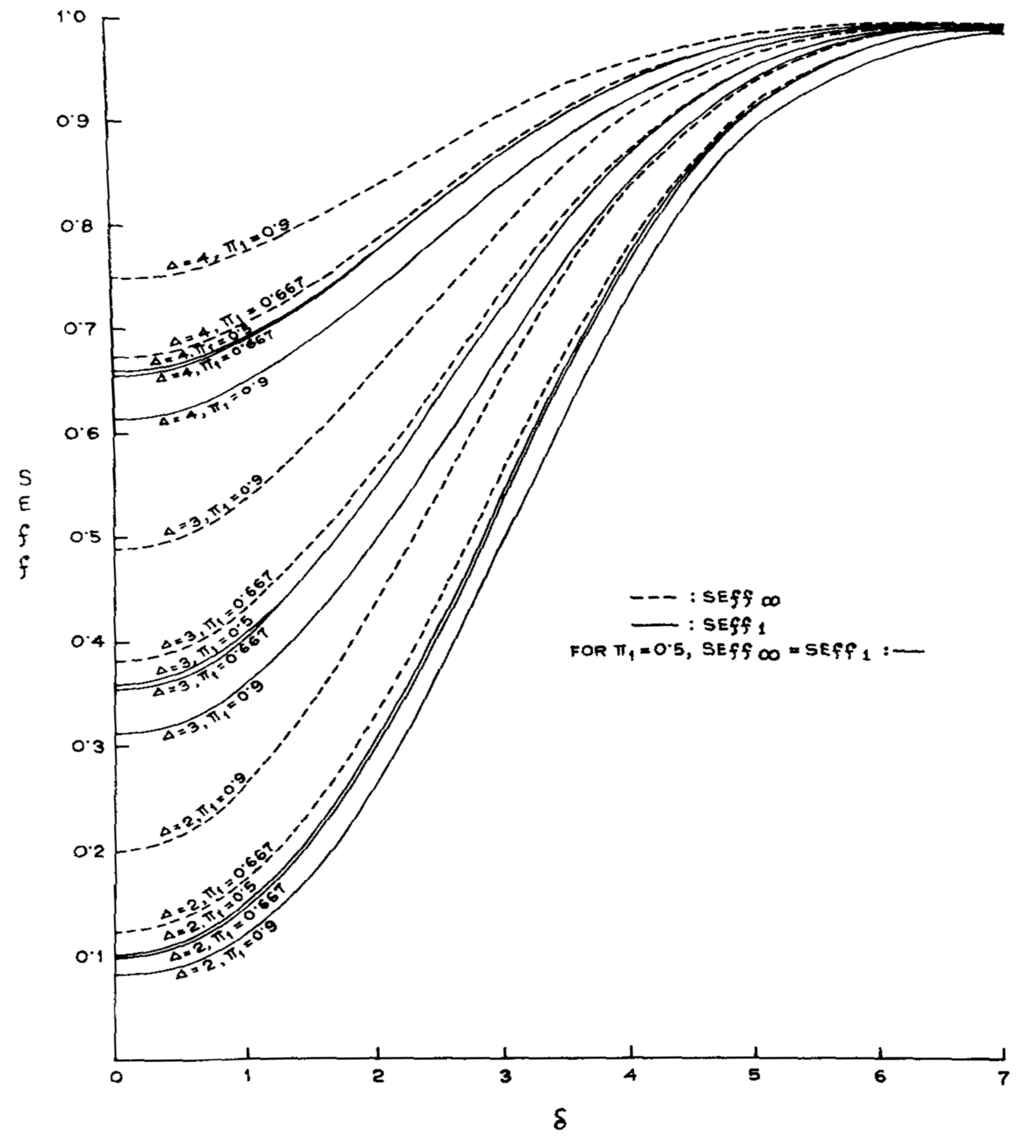

Fig. 1. Eff of *logistic-normal* supervisor for various values of parameters and supervision index.

4. $SEff_1$ decreases with values of π_1 away from $\frac{1}{2}$ and $SEff_\infty$ increases with the value of π_1 away from $\frac{1}{2}$; thus the situation of unequal prevalence rates of the two groups needs a larger number of features for the same distance between groups;

5. For $\delta = 0$, $SEff_1$ and $SEff_\infty$ coincide with the corresponding efficiency in O'Neill's[7] tables; clearly, $\delta = 0$ is the unsupervised case and hence the SEff formula is the same as O'Neill's formula for the unsupervised case.

The interpretation of these efficiencies is similar to that pointed out in our earlier paper.[4] For instance, when the feature distribution parameter values are $p = 1$, $\Delta = 4$, $\pi_1 = 0.667$ and the logistic-normal supervision parameter is $\delta = 3$, $SEff = 0.8630$; this means that for these feature distribution parameter values, about 86 stochastically (δ) supervised samples are equivalent to 100 perfectly supervised samples.

Titterington[5] has established a similarity between the two models for some parametric values on the basis of equality of two moments; he shows that the beta model with $m = 5, n = 3$ is similar to the logistic-normal model with $\eta = 0.5833$ and $\sigma = 0.6112$ resulting in $\delta = 2.2262$; this is done on the basis of equating moments of order $r_1 = 0.05$ and $r_2 = 0.10$. However, if the first and second moments are equated we get for the same beta parameters $\eta = 0.5754$ and $\sigma = 0.5754$ resulting in $\delta = 1.5736$, in which case the two models look quite different. The values of r_1 and r_2 needed to strike a similarity between the two models are different for different m and n. Hence it is not easy to convert results of the beta model into those of the logistic-normal model. For the above case when the two models are similar, the efficiencies are also fairly equal as shown in Table 2, where the efficiencies for the beta case are computed using the formula in our previous article.[4] However, the problem remains as to how to set up a correspondence between these two models. In the case of the logistic-normal model an *index of supervision* is naturally given by δ. But it is not clear as to what

Table 2. Comparison of asymptotic relative efficiencies of normal discrimination under logistic-normal and beta supervision models

$\pi_1 = 0.5$

Model	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$
$\delta = 1.5736$	0.2262	0.2262	0.4802	0.4802	0.7332	0.7332
$\delta = 2.2262$	0.3572	0.3572	0.5852	0.5852	0.7931	0.7931
$m = 5, n = 3$	0.3911	0.3911	0.5503	0.5503	0.7543	0.7543

$\pi_1 = 0.667$

Model	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$
$\delta = 1.5736$	0.2242	0.2497	0.4777	0.4997	0.7305	0.7436
$\delta = 2.2262$	0.3549	0.3796	0.5825	0.6013	0.7905	0.8019
$m = 5, n = 3$	0.3824	0.4015	0.5360	0.5654	0.7439	0.7648

$\pi_1 = 0.9$

Model	$\Delta = 2$		$\Delta = 3$		$\Delta = 4$	
	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$	$SEff_1$	$SEff_\infty$
$\delta = 1.5736$	0.1878	0.3532	0.4304	0.6000	0.6921	0.8056
$\delta = 2.2262$	0.3098	0.4874	0.5370	0.6889	0.7567	0.8565
$m = 5, n = 3$	0.3976	0.4294	0.4776	0.6341	0.6824	0.8180

a suitable single *index of supervision* should be in the case of the beta model; as pointed out in Krishnan and Nandy,[4] this index will be a suitably normalised version of $|m - n|$ and it is not clear what this normalisation should be. A correspondence between the logistic-normal model and the beta model may depend on the answer to this question. We are working on this problem.

SUMMARY

We consider the problem of discriminant analysis when the supervisor's classification is stochastic, and deal with the problem of efficiency of this supervision relative to perfect supervision. For this, we use the supervision model of Titterington,[5] who proposed a logistic-normal distribution for the supervisor's probability of classifying an object into a group; he had suggested this as an alternative to the beta model we had earlier proposed.[3,4] Under this model of supervision and for the case of two p-dimensional normal populations with a common covariance matrix, we derive formulae for Efron efficiency of stochastic supervision, which is an index of the amount of statistical information contained in the stochastic supervision *vis-à-vis* perfect supervision. We present a table and a chart of this efficiency for various values of parameters of the two p-dimensional normal populations and various levels of supervision. We interpret these values to show how the worth of a stochastically supervised sample depends on the supervision parameters and the parameters of the feature distributions. Stochastically supervised initial samples are quite useful unless the supervision parameters are such that it is very close to a completely unsupervised situation.

REFERENCES

1. U. A. Katre and T. Krishnan, Pattern recognition with imperfect supervision, *Pattern Recognition* **22**, 423–432 (1989).
2. T. Krishnan, Efficiency of learning with imperfect supervision, *Pattern Recognition* **21**, 183–188 (1988).
3. T. Krishnan and S. C. Nandy, Discriminant analysis with a stochastic supervisor, *Pattern Recognition* **20**, 379–384 (1987).
4. T. Krishnan and S. C. Nandy, Efficiency of discriminant analysis when initial samples are classified stochastically, *Pattern Recognition* **23**, 529–537 (1990).
5. D. M. Titterington, An alternative stochastic supervisor in discriminant analysis, *Pattern Recognition* **22**, 91–95 (1989).
6. B. Efron, The efficiency of logistic regression compared to normal discriminant analysis, *J. Am. Stat. Assoc.* **70**, 892–898 (1975).
7. T. J. O'Neill, Normal discrimination with unclassified observations, *J. Am. Stat. Assoc.* **73**, 821–826 (1978).

About the Author—T. KRISHNAN received his Master's degree in Mathematics from Madras University in 1958 and Master's degree in Statistics and the Ph.D. degree from the Indian Statistical Institute in 1965 and 1968, respectively. He has been with the Indian Statistical Institute since then and is now a Professor of Applied Statistics. He has held visiting positions in the University of Southampton and the University of Western Australia. His research interests are in Pattern Recognition, Biostatistics and Psychometry.

About the Author—SUBHAS CHANDRA NANDY received his Master's degree in Statistics from Calcutta University in 1982 and the Master of Technology degree in Computer Science from the Indian Statistical Institute in 1985. He has been with the Indian Statistical Institute since then and is now a Programmer. His research interests are in Pattern Recognition, VLSI and related areas.