When $H_r(\omega)$ is well-defined we have either $D_{sn}^r(\omega) = -\min\{D_s(\omega), D_n^r(\omega)\}$ or $D_{sn}^r(\omega) = -U(\omega)$, from (13). In the former case $H_r(\omega)(1 - H_r(\omega))$ is zero, and in the latter case $H_r(\omega)$ and thus $H_r(\omega)(1 - H_r(\omega))$ lies in $[0, 1]$, and $\mathrm{Re}\{D_{sn}^r(\omega)\} - \mathrm{Re}\{D_{sn}(\omega)\} \leqslant 0$. Thus when $H_r(\omega)$ is well-defined,

$$P(\omega) \geqslant H_r^2(\omega)[D_n^r(\omega) - D_n(\omega)]. \tag{A2}$$

Equation (4) does not yield a well-defined $H_r(\omega)$ when $D_s(\omega) = D_n^r(\omega) = -D_{sn}^r(\omega)$. In this case, using the inequality

$$\mathrm{Re}\{D_{sn}(\omega)\} \geqslant -\frac{D_s(\omega) + D_n(\omega)}{2}$$

(from $\mathrm{Re}\{D_{sn}(\omega)\} \geqslant -|D_{sn}(\omega)| \geqslant -\sqrt{D_s(\omega) D_n(\omega)} \geqslant -\frac{1}{2}[D_s(\omega) + D_n(\omega)]$) in (A1), we get for $0 \leqslant H_r(\omega) \leqslant 1$

$$P(\omega) \geqslant H_r(\omega)[D_n^r(\omega) - D_n(\omega)]. \tag{A3}$$

a) When $D_n^r(\omega)$ in Case B is given as $D_n^r(\omega) = f_k(\omega)$, from (12) and because $\min\{D_s(\omega), U(\omega)\} \leqslant f_k(\omega)$, we have $|D_{sn}^r(\omega)| = \min\{D_s(\omega), U(\omega)\}$.

Now note that if (4) gives a well-defined H_r, it is (using 13)

$$H_r(\omega) = \frac{D_s(\omega) - |D_{sn}^r(\omega)|}{D_s(\omega) - |D_{sn}^r(\omega)| + D_n^r(\omega) - |D_{sn}^r(\omega)|} \tag{A4}$$

which in this case becomes $H_r(\omega) = 1/(1 + k)$. When (4) does not define H_r, it is taken to be $1/(1 + k)^2$. Thus, using (A2) and (A3), we have

$$P(\omega) \geqslant \frac{1}{(1 + k)^2}[D_n^r(\omega) - D_n(\omega)].$$

b) For values of ω where $D_n^r(\omega) = L_n(\omega)$, we again have $|D_{sn}^r(\omega)| = \min\{D_s(\omega), U(\omega)\}$, and (A4) now gives $H_r(\omega) \leqslant 1/(1 + k)$, because here $D_n^r(\omega) > f_k(\omega)$. This gives, because $D_n^r(\omega) - D_n(\omega) \leqslant 0$, the result

$$P(\omega) \geqslant \frac{1}{(1 + k)^2}[D_n^r(\omega) - D_n(\omega)].$$

c) When $D_n^r(\omega) = U_n(\omega)$, note that we have $D_n^r(\omega) < kD_s(\omega) + (1 - k)|D_{sn}^r(\omega)|$, since here $U_n(\omega) < f_k(\omega) \leqslant D_s(\omega)$ and $|D_{sn}^r(\omega)| = \min\{U_n(\omega), \min\{D_s(\omega), U(\omega)\}\}$. Thus from (A4) $H_r(\omega) \geqslant 1/(1 + k)$, and because here $D_n^r(\omega) - D_n(\omega) \geqslant 0$, we get again

$$P(\omega) \geqslant \frac{1}{(1 + k)^2}[D_n^r(\omega) - D_n(\omega)].$$

For all ω we have shown that the above inequality is true; the result follows, by integrating, that

$$e(D^r, H_r) - e(D, H_r) \geqslant 0.$$

REFERENCES

[1] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171-185, Oct./Nov. 1977.

[2] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 521-526, June 1980.

[3] H. V. Poor and D. P. Looze, "Minimax state estimation for linear stochastic systems with noise uncertainty," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 902-906, Aug. 1981.

[4] K. S. Vastola and H. V. Poor, "Robust linear estimation of stationary discrete-time signals," *Proc. 1981 Conf. Inform. Sci. Syst.*, The Johns Hopkins University, pp. 512-516, Mar. 1981.

[5] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753-1758, 1965.

[6] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, 242-247, Mar. 1981.

[7] H. V. Poor, "Minimax linear smoothing for capacities," *Ann. Prob.*, vol. 10, pp. 504-507, May 1982.

[8] G. Moustakides and S. A. Kassam, "Robust Wiener filters for correlated signals and noise," *Proc. 1980 Conf. Inform. Sci. Syst.*, (Princeton University), pp. 308-313, Mar. 1980.

[9] G. Moustakides, "Robust Wiener filters for imprecise second-order statistics," M.S.E. thesis, Department of Systems Engineering, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, PA, Aug. 1980.

A Simple Suboptimum Estimator of Prior Probability in Mixtures

REZA JALALINASAB AND JAMES A. BUCKLEW, MEMBER, IEEE

***Abstract*—A simple relative frequency type estimator of the prior probability in a mixture of two known density functions is presented. Examples are given demonstrating the ease of design and implementation of this estimator structure.**

I. INTRODUCTION

Given that an unknown density function $f(x)$ is a mixture of two known density functions $f_1(x)$ and $f_2(x)$, where the prior probability is unknown, the problem is to estimate the prior probability on the basis of N statistically independent observations. The density $f(x)$ can be written as

$$f(x) = \pi f_1(x) + (1 - \pi) f_2(x), \tag{1}$$

where π the prior probability to be estimated, is assumed to be uniformly distributed on [0, 1].

This type of problem arises in pattern recognition problems where one is studying the distribution of observations belonging to individual populations and where the population mix is unknown. It also has application in biological and physical sciences. (See Choi [1], Makov and Smith [2], Davisson [3], Sakrison [4], Yakow [5], Blischke [6], and Makov [7] for examples and further references.)

The estimator is presented in Section III. It is a member of a class of estimators first presented by Boes [8]. In Sections IV and V we give three numerical examples, and we compare our results with a recursive estimation scheme due to Kazakos [9].

The calculations involved in designing Kazakos' estimator are moderately complicated. On the other hand, the estimator introduced in this paper is very simple to design. The examples in Section IV are intended to illustrate the simplicity of the estimator design. We emphasize that Kazakos' estimate always gives a lower variance, but in some applications one has a great number of samples and simplicity is more desirable.

II. DEVELOPMENT

Let f_1, f_2, π, and N independent samples $\{x_1, \cdots, x_n\}$ be as stated in the preceding section. Consider a set of the following form:

$$A = \left\{x: \frac{f_1(x)}{f_2(x)} \geqslant T\right\}, \tag{2}$$

where T is a threshold to be described later. Define

$$P(A) = \int_A f(x)\, dx \tag{3}$$

Manuscript received January 19, 1982; revised November 11, 1982.

R. Jalalinasab was with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI.

J. A. Bucklew is with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706.

and

$$P_i(A) = \int_A f_i(x)\, dx, \qquad i = 1,2. \tag{4}$$

For notational ease we will sometimes write simply P_i and suppress the argument A. Integrating both sides of (1) over the set A implies that $P(A) = \pi P_1(A) + (1 - \pi)P_2(A)$, or

$$\pi = \frac{P(A) - P_2(A)}{P_1(A) - P_2(A)},$$

where we assume that $P_1(A) \neq P_2(A)$. Define

$$\hat{P}(A) = \frac{1}{N} \sum_{i=1}^{N} I_A(x_i), \tag{5}$$

where I_A is the indicator function of the set A.

Our objective is to estimate π by

$$\hat{\pi}_A = \frac{\hat{P}(A) - P_2(A)}{P_1(A) - P_2(A)}$$

under the minimum mean square error criterion. We choose mean square error primarily since it is simple to compute. Note that the mean square error expression will be a function of the set A defined by (2), which itself is a function of T. Hence the expression of the mean square error will be a function of T. We will then choose that value of T which gives smallest mean square error. We note again that the estimator structure given above is not new [10]. Our purpose is to compare the performance of such a simple estimator structure with optimal schemes.

III. THE ESTIMATOR

As mentioned in the last section, we intend to find the expression for the mean square error and the optimum threshold (or the optimum set A) which minimizes it.

The mean square error is given by

$$\begin{aligned} E(\pi - \hat{\pi}_A)^2 &= E\left(\pi^2 + \pi_A^2 - 2\pi\hat{\pi}_A\right) \\ &= E(\pi^2) + E\left(\pi_A^2\right) - 2E(\pi\hat{\pi}_A). \end{aligned}$$

Now we compute the above expression term by term, under the assumption that π is a uniformly distributed random variable. We find that

$$E(\pi^2) = \int_0^1 \pi^2 f(\pi)\, d\pi = \frac{1}{3},$$

$$\begin{aligned} E\left[\hat{\pi}_A^2\right] &= \frac{1}{(P_1 - P_2)^2}\left[\frac{P_1 + P_2}{2N} + \frac{(N-1)\left(P_1^2 + P_2^2 + P_1P_2\right)}{3N}\right. \\ &\qquad \left. + P_2^2 - 2P_2\left(\frac{1}{2}\right)(P_1 + P_2)\right] \\ &= \frac{N-1}{3N} + \frac{1/2(P_1 + P_2) - P_1P_2}{N(P_1 - P_2)^2}, \end{aligned}$$

and

$$\begin{aligned} E[\pi\hat{\pi}_A] &= \frac{1}{P_1(A) - P_2(A)} \\ &\qquad \cdot \left[\frac{1}{6}\int_A (2f_1(x) + f_2(x))\, dx - \frac{P_2(A)}{2}\right] = \frac{1}{3}. \end{aligned}$$

Recall that the set A is chosen that $P_1(A) \neq P_2(A)$.

Next, we substitute the proper expressions for each term in the original equation to obtain the final expression for the mean square error:

$$\begin{aligned} E(\pi - \hat{\pi}_A)^2 &= \frac{1}{3} + \frac{N-1}{3N} + \frac{1/2(P_1 + P_2) - P_1P_2}{N(P_1 - P_2)^2} - 2\left(\frac{1}{3}\right) \\ &= -\frac{1}{3N} + \frac{\frac{1}{2}\int_{(A)} (f_1(x) + f_2(x))\, dx - \int_A f_1(x)\, dx \cdot \int_A f_2(x)\, dx}{N\left(\int_A f_1(x)\, dx - \int_A f_2(x)\, dx\right)^2}. \end{aligned}$$

Notice that as N goes to infinity, the mean square error approaches zero as expected.

All we need do now is to find the optimum set A (or the optimum threshold T), for which the above expression is minimized. To illustrate the method of finding this optimum set we give an example in the next section.

IV. EXAMPLE

As an example we assume that $f_1(x)$ and $f_2(x)$ are uniformly distributed on $[0,2]$ and $[1,4]$, respectively.

By computing the ratio of the two density functions $f_1(X)$ and $f_2(X)$ we notice that it takes on only three values:

$$\frac{f_1(x)}{f_2(x)} = \begin{cases} \infty, & 0 \leqslant x \leqslant 1; \\ 3/2, & 1 < x \leqslant 2; \\ 0, & 2 < x \leqslant 4. \end{cases}$$

Now we vary T and try to find the T for which the mean square error is minimized.

Suppose $T = 0$. This implies that $A = \{x\colon 0 \leqslant x \leqslant 4\}$ and

$$\int_A f_1(x)\, dx = \int_A f_2(x)\, dx,$$

which contradicts our original assumption $(P_1(A) \neq P_2(A))$. Therefore we exclude this choice of T.

If $0 < T \leqslant 3/2$ then $A = \{x\colon 0 \leqslant x \leqslant 2\}$, while

$$\begin{aligned} E(\pi - \hat{\pi}_A)^2 &= -\frac{1}{3N} + \frac{1/2(1 + 1/3) - (1)(1/3)}{N(1 - 1/3)^2} \\ &= \frac{1}{3N} + \frac{3}{4N} = \frac{5}{12N}, \end{aligned}$$

and therefore

$$NE(\pi - \hat{\pi}_A)^2 = \frac{5}{12N}.$$

If $3/2 < T < \infty$, then $A = \{x\colon 0 \leqslant x \leqslant 1\}$, and

$$E\left\{(\pi - \hat{\pi}_A)^2\right\} = -\frac{1}{3N} + \frac{1/2(1/2 + 0) - 1/2 \cdot 0}{N(1/2)^2} = \frac{2}{3N}.$$

The minimum mean square error is therefore $5/12N$ and this minimum occurs when $A = \{x\colon 0 \leqslant x \leqslant 2\}$.

V. COMPARISON OF THE VARIANCES

Kazakos' Estimate and Its Variance: Kazakos' estimate [9] is given by

$$\begin{aligned} \hat{\pi}_N &= \hat{\pi}_{N-1} + N^{-1}L(\hat{\pi}_{N-1})(f_1(x_N) - f_2(x_N)) \\ &\qquad \cdot \left[\hat{\pi}_{N-1}f_1(x_N) + (1 - \hat{\pi}_{N-1})f_2(x_N)\right]^{-1}, \end{aligned}$$

where $\hat{\pi}_N$ is the estimate of π after observing $x_1, x_2, \cdots, x_N$, and $L(\hat{\pi}_{N-1})$ is an adjustable gain function assumed to be real valued, positive, and bounded.

The variance of this estimate is (See Makov [6])

$$\operatorname{var}(\hat{\pi}_N|\pi) = L^2(\pi)D(\pi)[2L(\pi)D(\pi) - 1]^{-1}/N,$$

where

$$D(\pi) = \int (f_1(x) - f_2(x))^2 [\pi f_1(x) + (1-\pi) f_2(x)]^{-1} \, dx$$
$$= \int \frac{(f_1(x) - f_2(x))^2}{\pi f_1(x) + (1-\pi) f_2(x)} \, dx.$$

When the variance is minimized the optimum choice of $L(\pi)$ is given by $L(\pi) = D^{-1}(\pi)$. Substituting $L(\pi) = D^{-1}(\pi)$ for the variance, we get

$$\operatorname{var}(\hat{\pi}_N | \pi) = \frac{1}{N \int \frac{(f_1(x) - f_2(x))^2}{\pi f_1(x) + (1-\pi) f_2(x)} \, dx}.$$

We mention that this variance asymptotically meets the Cramer–Rao lower bound, and hence the estimate is efficient (2). Note that the above is the variance of the estimator given that π is known. Using the fact that $\hat{\pi}_A$ is unbiased it is simple to show that $E\{(\pi - \hat{\pi}_A)^2\} = E(\operatorname{var}(\hat{\pi}_A | \pi))$. To compare the performance of the estimators it seems reasonable to compare $NE\{(\pi - \hat{\pi}_A)^2\}$ with $NE(\operatorname{var}(\hat{\pi}_N | \pi))$.

Kazakos' variance for the example of the previous section is

$$N \operatorname{var}(\hat{\pi}_N | \pi)$$
$$= \frac{1}{\int_0^1 \frac{(1/2)^2}{\pi (1/2)} dx + \int_1^2 \frac{(1-6)^2}{\pi/2 + \frac{1-\pi}{3}} dx + \int_2^4 \frac{(1/3)^2}{\frac{1-\pi}{3}} dx}$$
$$= \frac{1}{1/2\pi + \frac{1}{6(\pi+2)} + \frac{2}{3(1-\pi)}}.$$

Taking the expected value of the above with respect to π yields

$$NE(\operatorname{var} \hat{\pi}_N | \pi) = E\left[\frac{1}{1/2 + \frac{1}{6(\pi+2)} + \frac{2}{3(1-\pi)}} \right] = 0.2804.$$

Note that the average variance of Kazakos' estimate is less than that of our estimate, with a difference of about 33 percent.

We note that we have done other calculations comparing the two estimators. If $f_1(x)$ is normal with zero mean and unit variance and $f_2(x)$ is normal with a mean of 1/2 and a variance of 1/4, then the Kazakos estimator is about 48 percent better. If $f_1(x)$ is normal with zero mean and unit variance and $f_2(x)$ is Rayleigh with parameter 1 then Kazakos' estimator is about 57 percent better.

VI. CONCLUSION

We have studied the estimation of the prior probability in a mixture of two known density functions on the basis of a set of N independent observations. A relative frequency estimator design was established using the mean square error criterion. Using three numerical examples, we compared the average variance of an efficient estimate with the average variance of our estimate. This comparison shows that there is a trade-off between an easily implementable estimator and an estimator with a low variance. Our estimator generally only requires threshold logic and a counter. Kazakos' estimate gives a low variance which asymptotically meets the Cramer–Rao lower bound, but the computations involved are more complicated.

REFERENCES

[1] K. Choi, "Estimate of the parameters of a finite mixture distributions," Univ. of Missouri, Tech. Rep. 18, 1966.

[2] U. E. Makov and A. E. Smith, "A quasi-Bayes unsupervised learning procedure for priors," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 761–764, Nov. 1977.

[3] L. D. Davisson, "Convergence probability bounds for stochastic approximation," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 680–685, Nov. 1970.

[4] D. Sakrison, "Stochastic approximation: A recursive method for solving regression problems," in *Advances in Communication Systems*, Vol. 2, A. V. Balakrishnan, Ed. New York: Academic, 1966, pp. 51–106.

[5] S. Yakowitz, "A consistent estimator for the identification of finite mixtures," *Ann. Math. Statist.*, vol. 40, pp. 1728–1735, Aug. 1969.

[6] W. R. Bliscke, "Estimating the parameters of mixtures of binomial distributions," *J. Amer. Statist. Ass.*, vol. 59, 510–528, 1964.

[7] U. E. Makov, "On the choice of gain functions in recursive estimation of prior probabilities," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 4, July 1980.

[8] D. C. Boes, "On the estimation of mixing distributions," *Ann. Math. Statist.*, vol. 37, pp. 177–188, Jan. 1966.

[9] D. Kazakos, "Recursive estimation of prior probabilities using a mixture," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 203–211, Mar. 1977.

[10] H. Robbins, "The empirical Bayes approach to testing statistical hypothesis," *Rev. Int. Statist. Inst.*, vol. 31, pp. 195–208, 1963.

ACKNOWLEDGMENT

The authors would like to thank G. L. Wise of the University of Texas and L. Devroye of McGill University for a discussion where the seed of the idea for this correspondence was planted.

Stochastic Reliability Functions for Failure Rates Derived from Gauss–Markov Processes

JOSEPH L. HIBEY, MEMBER, IEEE

***Abstract*—An extension of the well-known Cameron–Martin formula can be interpreted as the expectation of a stochastic reliability function applicable in those situations where nondecreasing failure rates are desired. This follows if the failure rate is modeled as the square of a Gauss–Markov process. We describe the methodology for the general vector case, and then specialize the results to the one-dimensional case so as to obtain an exact closed-form expression for the reliability function. Using the theory of recurrent and transient processes, we then show how the choice of a model parameter and the initial state influence reliability.**

I. INTRODUCTION

In the study of both hardware and software reliability, one is sometimes faced with having to choose a mathematical model for the underlying failure rate mechanism. The literature on this subject is replete with both deterministic and stochastic models. Some models of each are given by Komoriya *et al.* [9] in the context of software reliability for the nuclear power industry. The authors show that a stochastic model proposed by Littlewood (and Verrall) [12] results in a reliability function that is more conservative than any of the deterministic models considered. Since such conservatism is a highly desirable design objective, we are thereby motivated to study other stochastic models.

Manuscript received January 19, 1982; revised July 12, 1982. Part of this work was supported while the author was a faculty research participant at Argonne National Laboratory by the U.S. Department of Energy, Office of Energy Research, through its University/National Laboratory Cooperative Program. This work was presented (with some discussion of filtering) at the 20th Annual Allerton Conference on Communication, Control, and Computing, October 6–8, 1982, at the University of Illinois, Urbana-Champaign, IL 61801.

The author is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

0018-9448/83/0700-0621$01.00 ©1983 IEEE