Reducing Probability of Decision Error using Stochastic Resonance

Steven Kay*
Dept. of Electrical and Computer Engineering
University of Rhode Island
Kingston, RI 02881
401-874-5804 (voice) 401-782-6422 (fax)
kay@ele.uri.edu

James H. Michels*
JHM Technologies
P.O. Box 4142
Ithaca, NY 14852
(607) 257- 4081
jmichels@americu.net

Hao Chen*, Pramod K. Varshney*
Department of EECS
335 Link Hall
Syracuse University
Syracuse, NY 13244
(315) 443-4013 (Pramod)
varshney@syr.edu
hchen21@syr.edu

EDICS: SAS - STAT
Keywords: signal detection, modeling, pattern classification

April 6, 2006

Abstract

The problem of reducing the probability of decision error of an existing binary receiver that is suboptimal using the ideas of stochastic resonance is solved. The optimal probability density function of the random variable that should be added to the input is found to be a Dirac delta function, and hence the optimal random variable is a constant. The constant to be added depends upon the decision regions and the probability density functions under the two hypotheses, and is illustrated with an example. Also, an approximate procedure for the constant determination is derived for the mean-shifted binary hypothesis testing problem.

*This work was supported by the Air Force Office of Scientific Research under contract FA9550-05-C-0139.

1 Introduction

The phenomenon of stochastic resonance has garnered much attention [1–4]. In short, it asserts that many physical processes in nature can be modeled as a detector over which we normally have no control. However, the stimulus to the process or equivalently the input to the detector is a quantify over which we do exert some measure of control. For example, in human image perception it is well known that contrast enhancement aids recognition of objects seemingly "buried" within the image. Hence, it is of importance to understand how one can modify the input to enhance the decision process. In stochastic resonance, the input is modified by adding a random variable or more generally noise. Recently, some approaches to determine the optimal type of noise to be added to a data set to improve detection performance have been derived [5]. In this paper we address the similar hypothesis testing problem of attempting to decide between two hypotheses but where the performance criterion is the probability of decision error. In [5] a Neyman-Pearson criterion is utilized.

We consider the problem of deciding between two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ that can occur with a priori probabilities $P[\mathcal{H}_0] = \pi_0$ and $P[\mathcal{H}_1] = \pi_1 = 1 - \pi_0$, respectively. Our criterion for performance will be probability of error P_e, although the derivation is easily modified to minimize the Bayes' risk by assigning costs associated with each decision [6]. It is assumed that *the decision regions have already been specified, that they are not optimal in terms of minimizing* P_e, and that a single data sample x is used to make a decision. The already specified decision regions may be arbitrary and hence our solution encompasses such regions as if one would decide $\mathcal{H}_1$ if $x > a$ or $|x| < a$ as examples. The single sample is usually a test statistic, i.e., a function of a set of observations, which is a common procedure for decision making. To improve the performance a "noise sample" c is added to form $y = x + c$ prior to decision making. We allow c to be a *random variable* and determine the PDF of c that will yield the minimum P_e. It is proven next that the optimal PDF is a Dirac delta function, which leads to the conclusion that *the optimal random variable to be added is a degenerate one, i.e., a constant.*

2 Optimal PDF of Additive Noise Sample

To write the probability of error for the original problem we define the decision rule (also called the test function or critical region indicator function) as

$$\phi(x) = \begin{cases} 0 & \text{decide } \mathcal{H}_0 \\ 1 & \text{decide } \mathcal{H}_1. \end{cases}$$

Then, we have

$$P_e = P[\text{decide } \mathcal{H}_1 | \mathcal{H}_0] P[\mathcal{H}_0] + P[\text{decide } \mathcal{H}_0 | \mathcal{H}_1] P[\mathcal{H}_1]$$

$$
\begin{aligned}
&= P[\phi(x)=1|\mathcal{H}_0]\pi_0 + P[\phi(x)=0|\mathcal{H}_1]\pi_1 \\
&= \pi_0 \int_{-\infty}^{\infty} \phi(x) p_0^X(x)dx + \pi_1 \int_{-\infty}^{\infty} (1-\phi(x)) p_1^X(x)dx
\end{aligned}
$$

where $p_0^X(x), p_1^X(x)$ are the probability density functions (PDFs) under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. This can be rewritten as

$$P_e = \pi_1 + \int_{-\infty}^{\infty} \phi(x)(\pi_0 p_0^X(x) - \pi_1 p_1^X(x))dx.$$

Now assume that we modify x by adding c so that the test statistic becomes $y = x + c$, where c is a random variable independent of x, and whose PDF is $p_C(c)$. Since the identical decision rule is to be used, we have

$$P_e = \pi_1 + \int_{-\infty}^{\infty} \phi(y)(\pi_0 p_0^Y(y) - \pi_1 p_1^Y(y))dy.$$

But

$$
\begin{aligned}
p_0^Y(y) &= \int_{-\infty}^{\infty} p_0^X(y-c)p_C(c)dc \\
p_1^Y(y) &= \int_{-\infty}^{\infty} p_1^X(y-c)p_C(c)dc.
\end{aligned}
$$

We have then that

$$
\begin{aligned}
P_e &= \pi_1 + \int_{-\infty}^{\infty} \phi(y) \left[\pi_0 \int_{-\infty}^{\infty} p_0^X(y-c)p_C(c)dc - \pi_1 \int_{-\infty}^{\infty} p_1^X(y-c)p_C(c)dc\right] dy \\
&= \pi_1 + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \phi(y) \left(\pi_0 p_0^X(y-c) - \pi_1 p_1^X(y-c)\right) dy\right] p_C(c)dc \\
&= \pi_1 + E_C \left[\int_{-\infty}^{\infty} \phi(y) \left(\pi_0 p_0^X(y-c) - \pi_1 p_1^X(y-c)\right) dy\right]
\end{aligned}
$$

where E_C denotes expected value. Hence, we wish to choose $p_C(c)$ so that the slightly more convenient form

$$J(p_C) = E_C \left[\int_{-\infty}^{\infty} \phi(y) \left(\pi_1 p_1^X(y-c) - \pi_0 p_0^X(y-c)\right) dy\right] \tag{1}$$

is *maximized*. This is done in the next section. We will see that the random variable C may be chosen as a constant and therefore we need only maximize the expression within the brackets of (1) over a constant c. But this is equivalent to shifting $\phi(u)$, the decision region function by $-c$. Hence, the solution effectively shifts the decision region by a constant. This suggests that another means for improving performance is to transform the decision region using a *nonlinear* transformation (instead of the simple shift). It can be done by transforming the data sample x using a nonlinear transformation g as $g(x)$. A future paper will address this alternative and more general approach.

3 Derivation of Optimal PDF for C

It is well known that $E_C[g(C)]$ is maximized by placing all the probability mass at the value of c for which $g(c)$ is maximized. We assume that the function $g(c)$ has at least one point at which a maximum is attained. Calling this point c_0 the optimal PDF is then $p_C(c) = \delta(c - c_0)$, where

$$c_0 = \arg_c \max g(c)$$

or

$$c_0 = \arg_c \max \int_{-\infty}^{\infty} \phi(y) \left(\pi_1 p_1^X(y-c) - \pi_0 p_0^X(y-c)\right) dy.$$

A slightly more convenient form for $g(c)$ is obtained by letting $u = y - c$ so that

$$g(c) = \int_{-\infty}^{\infty} \phi(u+c) \left(\pi_1 p_1^X(u) - \pi_0 p_0^X(u)\right) du \tag{2}$$

which is recognized as a *correlation* between $\phi(u)$ and $\pi_1 p_1^X(u) - \pi_0 p_0^X(u)$. In summary, *we should add the constant c to x, where c is the value that maximizes the correlation given in (2).* Since the decision function $\phi(x)$ in (2) is completely general, the optimal solution is valid for a given binary decision rule with *any decision region*. For example, if the original decision rule were to decide $\mathcal{H}_1$ if $x > a$, then we would use $\phi(u) = 1$ for $u > a$ and zero otherwise in (2). If it were to decide $\mathcal{H}_1$ if $|x| < a$, then $\phi(u) = 1$ for $|u| < a$, and zero otherwise, then we would use $\phi(u) = 1$ or $|u| < a$, and zero otherwise in (2). (Note that if $\phi(u) = 1$ for $\pi_1 p_1^X(u) - \pi_0 p_0^X(u) > 0$ and zero otherwise, then $g(c)$ is maximized for $c = 0$. This is because in this case the decision rule $\phi(u)$ is optimal.) In the next section we solve this for a given example.

4 The Gaussian Mixture Example

We now consider the problem described in [4] but instead choose the probability of error criterion. The problem is to decide between $p_0^X(x)$ and $p_1^X(x) = p_0^X(x - A)$, where $A > 0$ is a DC level that is known and the noise PDF is the Gaussian or normal mixture

$$p_0^X(x) = \frac{1}{2}\mathcal{N}(x; \mu, \sigma^2) + \frac{1}{2}\mathcal{N}(x; -\mu, \sigma^2) \tag{3}$$

where

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right].$$

The original decision rule is to choose $\mathcal{H}_1$ if $x > 0$ so that $\phi(x) = u_s(x)$, where $u_s(x)$ is the unit step function. Additionally, we assume equal a priori probabilities so that $\pi_0 = \pi_1 = 1/2$. As a result, we have from (2) that

$$\begin{aligned} g(c) &= \frac{1}{2}\int_{-\infty}^{\infty} u_s(u+c)(p_1^X(u) - p_0^X(u))du \\ &= \frac{1}{2}\int_{-c}^{\infty} (p_1^X(u) - p_0^X(u))du \\ &= \frac{1}{2}[(1 - F_1(-c)) - (1 - F_0(-c))] \\ &= \frac{1}{2}(F_0(-c) - F_1(-c)) \end{aligned}$$

where F_i is the cumulative distribution function of x under the hypothesis $\mathcal{H}_i$. For our problem we have that $p_1^X(x) = p_0^X(x - A)$ and so $F_1(x) = F_0(x - A)$. Thus,

$$g(c) = \frac{1}{2}(F_0(-c) - F_0(-c - A))$$

and differentiating and setting equal to zero produces

$$p_0^X(-c) = p_0^X(-c - A)$$

or equivalently since $p_0^X(x)$ is even, we have the general requirement

$$p_0^X(c) = p_0^X(c + A). \tag{4}$$

Using (3) produces

$$\phi(c; \mu, \sigma^2) + \phi(c; -\mu, \sigma^2) = \phi(c + A; \mu, \sigma^2) + \phi(c + A; -\mu, \sigma^2)$$

which upon simplification yields the equation

$$\begin{aligned} \exp(\mu c/\sigma^2) + \exp(-\mu c/\sigma^2) &= \exp[-c(A - \mu)/\sigma^2 - A^2/(2\sigma^2) + \mu A/\sigma^2] \\ &\quad + \exp[-c(A + \mu)/\sigma^2 - A^2/(2\sigma^2) - \mu A/\sigma^2]. \end{aligned}$$

For $\mu = 3$, $\sigma^2 = 1$, and $A = 1$ we have

$$\exp(3c) + \exp(-3c) = \exp(2c + 5/2) + \exp(-4c - 7/2).$$

The exact value of c found through a numerical search is $c = 2.5000$, which could also be found by ignoring the terms $\exp(-3c)$ and $\exp(-4c - 7/2)$ since these are nearly zero for this value of c. Another solution is found by ignoring the other set of terms to yield $c = -3.5$. Note that either of these choices causes the PDFs of $x + c$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ to cross at the origin. See Figures 1 and 2. If we did not have the right-most Gaussian mode, then the choice of $c = 2.5$ would result in a maximum likelihood (ML) receiver, which is optimum [6]. This is because a maximum likelihood receiver chooses the hypothesis whose PDF value is larger. In our case the *fixed* decision regions are $R_1 = \{x : x > 0\}$ for $\mathcal{H}_1$ and $R_0 = \{x : x < 0\}$ for $\mathcal{H}_0$ as shown in Figure 1. These decision regions are not optimal. The optimal ML decision regions are indicated in Figure 1 as R_0^* and R_1^*. Therefore, the region in x for which $R_i \neq R_i^*$, which corresponds to the dark PDF lines, will result in incorrect decisions. By the addition of c, however, the extent of this incorrect decision region is reduced, as indicated in Figure 2.

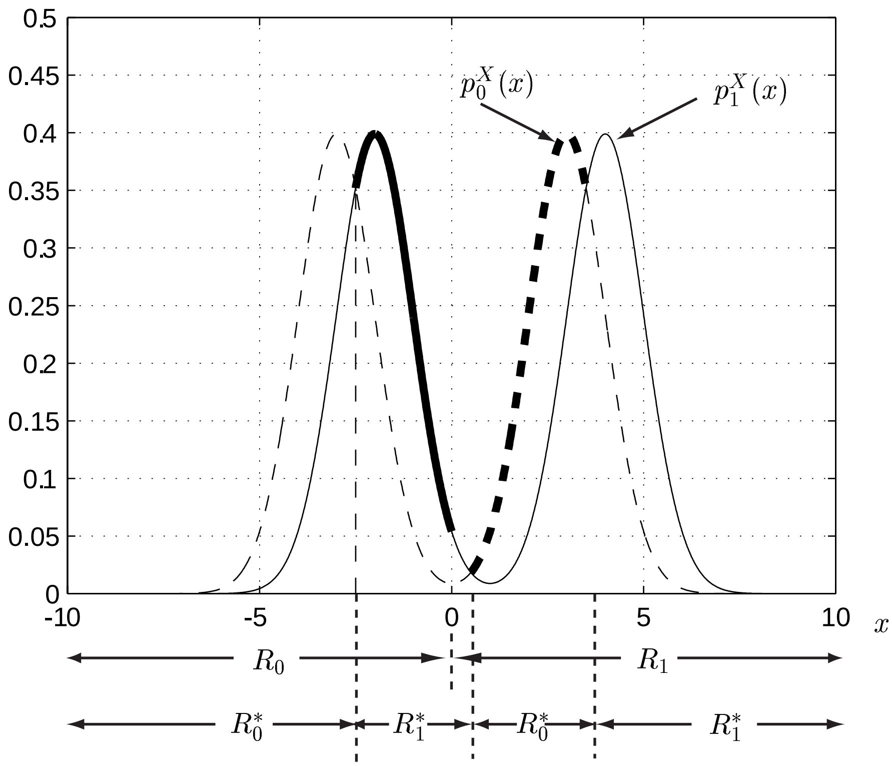

Figure 1: Original PDFs. The left-most PDF modes cross at $x = -2.5$, which is indicated by the dashed vertical line. The *fixed* decision regions are indicated by R_i while the optimal ML decision regions are indicated by R_i^*.

It is instructive to also plot the probability of error versus c or equivalently the probability of correct decision $P_c = 1 - P_e$ versus c. This is shown in Figure 3. Note that as expected the probability of a correct decision is maximized at $c = 2.5$ and also at $c = -3.5$. This type of curve is normally associated with the phenomenon of stochastic resonance, although here we see that it is *not unimodal.* This result is unlike that reported in [1–3] and so debunks the common assumption that adding too much noise will degrade performance. The latter is only true if the performance curve is unimodal.

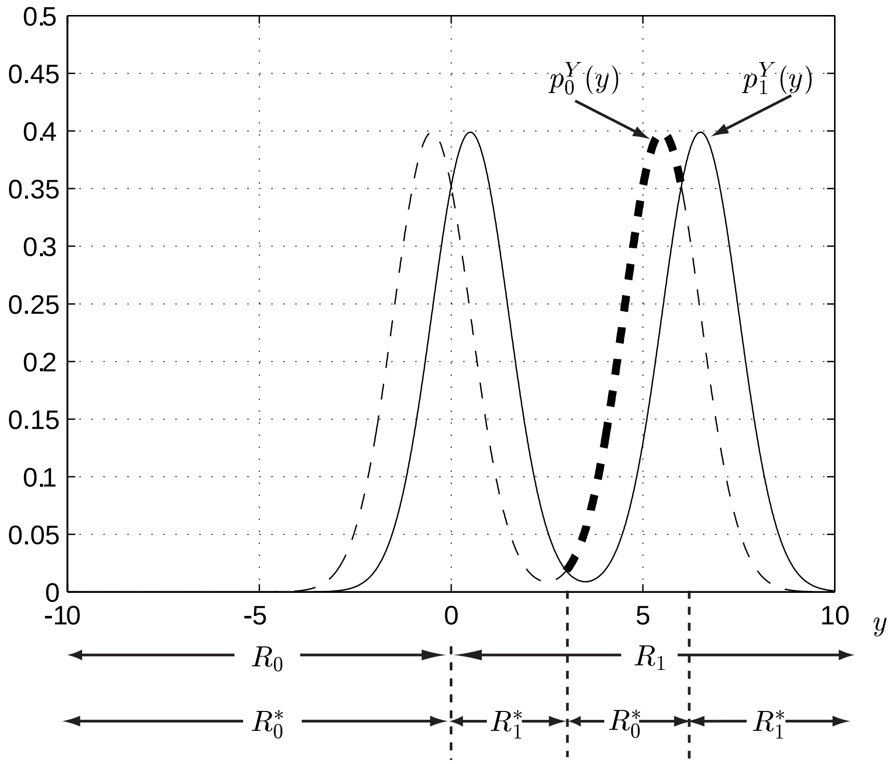

Figure 2: PDFs after $c = 2.5$ is added to x. The *fixed* decision regions are indicated by R_i while the optimal ML decision regions are indicated by R_i^*

5 A Simple Approximation

In some cases we can simplify the determination of c rather than having to solve (4). Consider again the case when $\pi_0 = \pi_1 = 1/2$ and $p_1^X(x) = p_0^X(x - A)$. This is an equal a priori probability and mean-shifted decision problem. Furthermore, assume that $A > 0$ and A is small. Finally, assume that the decision is to choose $\mathcal{H}_1$ if $y > 0$. Then, we have that in (2)

$$\begin{aligned}
\pi_1 p_1^X(u) - \pi_0 p_0^X(u) &= \frac{1}{2}(p_0^X(u - A) - p_0^X(u)) \\
&\approx \frac{1}{2}\left(p_0^X(u) - \frac{dp_0^X(u)}{du}A - p_0^X(u)\right) \\
&= -\frac{1}{2}\frac{dp_0^X(u)}{du}A
\end{aligned}$$

where we have used a first-order Taylor expansion in A about $A = 0$. Therefore,

$$g(c) \approx \int_{-\infty}^{\infty} \phi(u + c)\left[-\frac{1}{2}\frac{dp_0^X(u)}{du}A\right] du$$

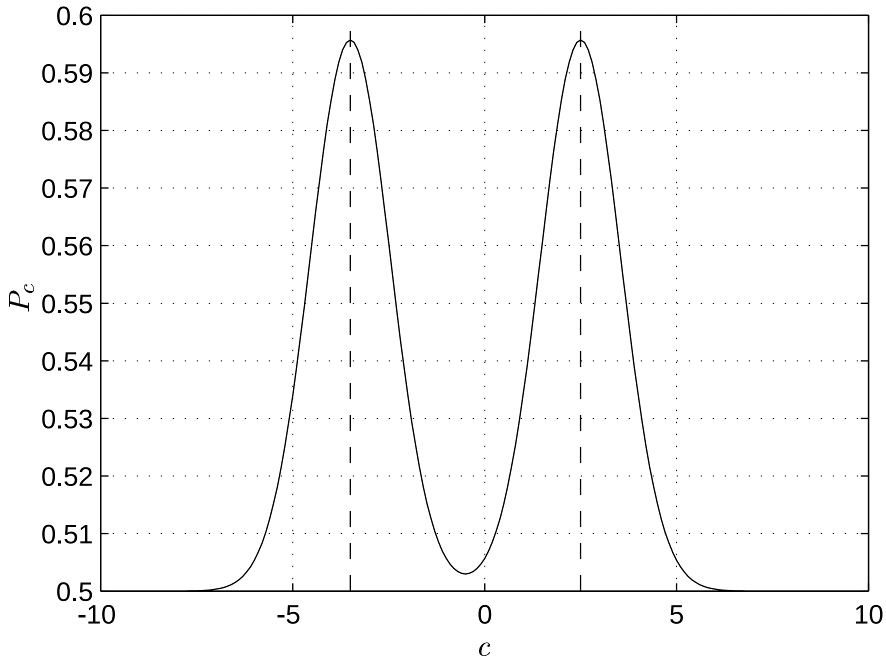

Figure 3: Probability of correct decision versus the value of the constant c to be added to data sample. The dashed lines are at $c = -3.5$ and $c = 2.5$.

$$
\begin{aligned}
&= -\frac{1}{2}A\int_{-c}^{\infty}\frac{dp_0^X(u)}{du}du \\
&= -\frac{1}{2}A\left[p_0^X(\infty) - p_0^X(-c)\right] \\
&= \frac{1}{2}Ap_0^X(-c)
\end{aligned}
$$

by noting that the PDF must converge to zero as its argument goes to infinity. Hence, to maximize $g(c)$ we need only find the location of the maximum of $p_0^X(x)$. In practice, since knowledge of the PDFs is usually lacking, this result will simplify the required knowledge necessary for implementation. Armed with actual data one should then be able to estimate the most probable value of the PDF under $\mathcal{H}_0$. Then, the optimal value of c is the negative of this. For example, in the Gaussian mixture example, we have maxima of $p_0^X(x)$ at approximately $x = \pm 3$ so that the optimal value of c is also ± 3. This is very close to our previous results of $c = -3.5$ and $c = 2.5$ and will be exact as $A \to 0$.

References

1. R. Benzit, A. Sutera, A. Vulpiani, "The Mechanism of Stochastic Resonance", *Journal of Physics A*, Vol 14, pp. L453–L457, 1981.

2. L. Gammaitoni, P. Hanggi, P. Jung, F. Marchesoni, "Stochastic Resonance", *Reviews of Modern Physics*, Vol. 70, pp. 223-287, Jan. 1998.

3. S. Mitaim, B. Kosko, "Adaptive Stochastic Resonance", *Proc. of the IEEE*, Vol. 86, pp. 2152–2183, 1998.

4. S. Kay, "Can Detectability be Improved by Adding Noise?", *IEEE Signal Processing Letters*, Jan. 2000.

5. H. Chen, P. Varshney, J. Michels, S. Kay, "Optimal Stochastic Resonance", to be presented at *2006 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Toulouse, France.

6. S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Upper Saddle River, NJ, 1998.