

Confusion and clairvoyance: some remarks on the composite hypothesis testing problem

James Theiler

Los Alamos National Laboratory

ABSTRACT

This paper discusses issues related to the inherent ambiguity of the composite hypothesis testing problem, a problem that is central to the detection of target signals in cluttered backgrounds. In particular, the paper examines the recently proposed method of continuum fusion (which, because it combines an ensemble of clairvoyant detectors, might also be called *clairvoyant fusion*), and its relationship to other strategies for composite hypothesis testing. A specific example involving the affine subspace model adds to the confusion by illustrating irreconcilable differences between Bayesian and non-Bayesian approaches to target detection.

Keywords: hypothesis testing, composite hypothesis testing, statistical data analysis, continuum fusion, clairvoyant fusion, target detection, multispectral, hyperspectral

“Joe Lightcap was not a philosopher; he took ideas seriously.”
—Edward Abbey, *The Fool’s Progress*

1. INTRODUCTION

The composite hypothesis testing problem is one of the great unsolved problems of statistics – but it is not unsolved because it is particularly hard; it is unsolved because it is fundamentally ambiguous. It is also enormously useful: it lies at the core of what it means to do science, and provides an excellent framework for doing target detection in multispectral imagery. While some of the ambiguities that arise in composite hypothesis testing are naturally modeled with probability, others are more difficult. These difficulties are in some ways more philosophical than they are mathematical, but we are not philosophers, so we will do what we can with mathematics.

For *simple* hypothesis testing, the aim is to distinguish which of two hypotheses is most consistent with observed data. This problem is straightforward, and unambiguously optimal solutions can be expressed in terms of likelihood ratios.

It gets confusing (or composite) when the aim instead is to distinguish between two *families* of hypotheses. The clairvoyant solution chooses a single member from each family and then uses the the ratio of their likelihoods. Although the clairvoyant solution isn’t very useful by itself (since, by the very statement of the problem, there’s no way to know which member to choose), it provides a valuable building block for constructing more effective solutions to the composite hypothesis testing problem.

The generalized likelihood ratio (GLR) has long been the workhorse solution for composite hypothesis testing problems, and for good reason: it is straightforward, unambiguous, and quite general. But it is not the only solution, and (except for a very few cases) it is not the optimal solution.

Recently, a new mathematical approach, called *continuum fusion*, was suggested as a way to address the composite hypothesis testing problem.^{1–10} This new approach generalizes the GLR (thus making it an “even more generalized” likelihood ratio), yet retains its min-max approach. This paper will attempt to make some mathematical observations about continuum fusion, but its main contribution will be to suggest a colorful new name – “clairvoyant fusion” – which shares the same acronym (CF) and more incisively describes the fusion process.

E-mail: jt@lanl.gov

Our interest is in detecting targets that are small or weak or rare (in short, that are difficult to detect) in backgrounds that are large and cluttered. In the statistical hypothesis testing framework, two hypotheses are considered: the “null” hypothesis \mathcal{H}_o is that there is no target, while the “alternative” \mathcal{H}_1 says that a target is present. From a measurement \mathbf{x} , the task is to choose between \mathcal{H}_o and \mathcal{H}_1 . More generally, a *detector* is a binary function[†] $\mathcal{B}(\mathbf{x})$ whose value (0 or 1) specifies the choice of hypothesis for each \mathbf{x} . There are two kinds of errors that can be made: if there is no target but $\mathcal{B}(\mathbf{x}) = 1$, that is a false alarm; if a target is present but $\mathcal{B}(\mathbf{x}) = 0$, that is a missed detection.

Many detectors are implemented in terms of a real-valued function $\mathcal{T}(\mathbf{x})$ which characterizes the “targetness” corresponding the measurement \mathbf{x} . A larger $\mathcal{T}(\mathbf{x})$ generally indicates more confidence that there is a target where \mathbf{x} was measured. By comparing this function to a threshold, one obtains a detector. This can be written $\mathcal{B}(\mathbf{x}) = \{\mathcal{T}(\mathbf{x}) \leq \lambda\}$, where the ‘ \leq ’ symbol corresponds to the notion that when $\mathcal{T}(\mathbf{x}) < \lambda$, then $\mathcal{B}(\mathbf{x}) = 0$ and \mathbf{x} is declared a background element; and when $\mathcal{T}(\mathbf{x}) > \lambda$, then $\mathcal{B}(\mathbf{x}) = 1$ and \mathbf{x} is declared a target.[‡]

2. SIMPLE HYPOTHESIS TESTING

When both the target and the background are drawn from *known* distributions – call them $p_t(\mathbf{x})$ and $p_b(\mathbf{x})$, respectively – the analysis is straightforward. The false alarm rate P_{fa} corresponds to the fraction of background pixels which are declared targets, and the detection rate is the fraction of actual target pixels which are declared targets; that is:

$$P_{fa} = \int \mathcal{B}(\mathbf{x})p_b(\mathbf{x}) d\mathbf{x}; \text{ and } P_d = \int \mathcal{B}(\mathbf{x})p_t(\mathbf{x}) d\mathbf{x}. \quad (1)$$

In comparing two detectors, we can say that one is more powerful than the other if its P_d is larger than the other’s P_{fa} , while its P_{fa} is at least as small as the other’s P_{fa} .

For the simple hypothesis testing problem, there is an optimal detector that is more powerful than all others; this detector is given by the likelihood ratio (LR).

$$\mathcal{T}_{LR}(\mathbf{x}) = \frac{p_t(\mathbf{x})}{p_b(\mathbf{x})} \leq \lambda. \quad (2)$$

That the likelihood ratio is optimal is known as the Neyman-Pearson theorem; Kay’s book¹² gives a nice demonstration of this optimality, while Theorem 3.2.1 in Lehmann and Romano’s book¹¹ provides a more formal proof.

Generically, this optimum is unique; that is, if two detectors are both optimal, they are generally identical. Exceptions can arise if the set $\{\mathbf{x} : \mathcal{T}_{LR}(\mathbf{x}) = \lambda\}$ has nonzero measure, but we will (in most cases, safely) assume that this never happens.

That this simple case can be optimized exactly only means the the uncertainty is minimized, not that it is eliminated. Given \mathbf{x} , we can make an unambiguously “best” guess, but we cannot be certain that it is correct.

3. COMPOSITE HYPOTHESIS TESTING

In the composite hypothesis testing problem, we are not only uncertain (probability can take care of that), we are confused. We do not have a single target distribution and a single background distribution. Instead, we have families of distributions for the target and/or background. We parameterize these families with $\theta_t \in \Theta_t$ and $\theta_b \in \Theta_b$ respectively. That is, $p_t(\mathbf{x}; \theta_t)$ is the distribution on \mathbf{x} when the target is present, and $p_b(\mathbf{x}; \theta_b)$ is the distribution when the target is absent. The problem is that the parameters, θ_t and/or θ_b , upon which the target and/or background distributions depend, are themselves unknown.

[†]In more sophisticated treatments,¹¹ the function is real-valued and varies from 0 to 1; it is treated as the probability of *labeling* \mathbf{x} a target. This is not to be confused with the probability of \mathbf{x} *being* a target. Generically, and for the examples in this paper, the set of \mathbf{x} for which the detector is non-deterministic is of measure zero, so the more sophisticated treatment is not necessary.

[‡]The case when $\mathcal{T}(\mathbf{x}) = \lambda$ provides yet another source of confusion, but it can be dealt with in an elegant and consistent way that is left to the professionals¹¹ (and hinted at in the previous footnote).

For any given (θ_t, θ_b) , we can characterize the performance of detectors using the definitions of false alarm rate and detection rate in Eq. (1), but those rates will depend on the unknown parameters.

$$P_{fa}(\theta_b) = \int \mathcal{B}(\mathbf{x}) p_b(\mathbf{x}, \theta_b) d\mathbf{x}; \text{ and } P_d(\theta_t) = \int \mathcal{B}(\mathbf{x}) p_t(\mathbf{x}, \theta_t) d\mathbf{x}, \quad (3)$$

For given parameter values (θ_t, θ_b) , we can use the definition in the previous section to say that one detector is more powerful than another. A stronger statement can be made if this comparison holds for all parameter values. In particular, one detector is said to be *uniformly* more powerful than another if it is at least as powerful as the other for every parameter value, and strictly more powerful for at least one parameter value. Regardless of what parameters (θ_t, θ_b) end up being appropriate, a uniformly more powerful detector is unambiguously preferable. The holy grail of composite hypothesis testing is to find the detector that is *uniformly most powerful* (UMP).^{11,12}

In many practical situations, however, no such UMP detector even exists. In these situations, when UMP is too much to ask for, we can make a more modest request, and ask of a detector that it at least is not uniformly less powerful than some other detector. A detector is *admissible* if it is not dominated by any other detectors; that is to say, there is no detector that is uniformly more powerful than an admissible detector.¹¹ For a given problem, there can be many distinct admissible detectors, no one of which is unambiguously superior to (*i.e.*, uniformly more powerful than) any other one. Which is “best” will depend on the operational requirements of the problem at hand, but in searching for this best detector, it makes sense to restrict the search to *admissible* detectors.

3.1. Clairvoyant detector

In the unrealistic case where one is given[†] the values of θ_t and θ_b , one can write the optimal detector in terms of the likelihood ratio:

$$\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) = \frac{p_t(\mathbf{x}; \theta_t)}{p_b(\mathbf{x}; \theta_b)} \leq \lambda. \quad (4)$$

This special detector, which depends on *a priori* knowledge of the parameter values, is called the “clairvoyant” detector.¹²

For composite hypothesis testing, these parameters are by definition *not* known. Nonetheless, we will find the clairvoyant detector to be a useful concept for building detectors that one could actually use in a composite hypothesis testing problem. Also, we find that some clairvoyant detectors, even though they are optimal only for only for a specific (θ_t, θ_b) , are nonetheless reasonable over the whole range of Θ_t and Θ_b .

As a general rule, specific clairvoyant detectors are *admissible*. Since the clairvoyant detector is optimal at its specific value of (θ_t, θ_b) , any candidate for a uniformly more powerful detector would have to be at least equally in power to the clairvoyant detector at the specific (θ_t, θ_b) , but since optimal detectors are generically unique, the other detector would then be identical to the clairvoyant detector.

In particular, if a composite hypothesis testing problem admits a UMP detector, then it will be equivalent to each of the clairvoyant detectors, which furthermore means that the clairvoyant detectors will all be equivalent to each other.

Another way to say that is: if two clairvoyant detectors are distinct, then the composite hypothesis testing problem does not admit a UMP solution. This is usually the situation with composite hypothesis testing.

3.2. Bayesian likelihood ratio

The Bayesian likelihood ratio (BLR)¹²⁻¹⁴ takes the philosophical position that the confusion in θ_t and θ_b in fact *can* be modeled with probability.

[†]*e.g.*, by some all-knowing crystal ball

The BLR employs priors for the θ_t and θ_b parameters; call them $\pi_t(\theta_t)$ and $\pi_b(\theta_b)$. It uses those priors to “marginalize out” θ_t and θ_b and to obtain posterior probabilities that do not involve these parameters. Specifically:

$$p_t^{\text{Bayes}}(\mathbf{x}) = \int_{\Theta_t} p_t(\mathbf{x}; \theta_t) \pi_t(\theta_t) d\theta_t, \quad (5)$$

$$p_b^{\text{Bayes}}(\mathbf{x}) = \int_{\Theta_b} p_b(\mathbf{x}; \theta_b) \pi_b(\theta_b) d\theta_b. \quad (6)$$

In terms of these posterior probabilities, we can employ a standard likelihood ratio:

$$\mathcal{T}_{\text{BLR}}(\mathbf{x}) = \frac{p_t^{\text{Bayes}}(\mathbf{x})}{p_b^{\text{Bayes}}(\mathbf{x})} = \frac{\int_{\Theta_t} p_t(\mathbf{x}; \theta_t) \pi_t(\theta_t) d\theta_t}{\int_{\Theta_b} p_b(\mathbf{x}; \theta_b) \pi_b(\theta_b) d\theta_b} \leq \lambda. \quad (7)$$

As an aside, we remark that the priors on θ_t and θ_b need not be independent. Given a prior on both of the parameters, $\pi(\theta_t, \theta_b)$, we use $\pi_t(\theta_t) = \int_{\Theta_b} \pi(\theta_t, \theta_b) d\theta_b$ and similarly for $\pi_b(\theta_b)$ in Eq. (7).

3.3. Generalized likelihood ratio (GLR) test

The most common non-Bayesian approach (and probably the most common approach, period) is the generalized likelihood ratio (GLR).¹²

$$\mathcal{T}_{\text{GLR}}(\mathbf{x}) = \frac{\max_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t)}{\max_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b)} \leq \lambda. \quad (8)$$

One interpretation of the GLR is that θ_t and θ_b are effectively re-estimated from each data element \mathbf{x} using maximum likelihood:

$$\hat{\theta}_t(\mathbf{x}) = \operatorname{argmax}_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t) \quad (9)$$

$$\hat{\theta}_b(\mathbf{x}) = \operatorname{argmax}_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b) \quad (10)$$

and then

$$\mathcal{T}_{\text{GLR}}(\mathbf{x}) = \mathcal{T}(\mathbf{x}; \hat{\theta}_t(\mathbf{x}), \hat{\theta}_b(\mathbf{x})) = \frac{p_t(\mathbf{x}; \hat{\theta}_t(\mathbf{x}))}{p_b(\mathbf{x}; \hat{\theta}_b(\mathbf{x}))} \leq \lambda. \quad (11)$$

Thus the GLR is like the clairvoyant detector except that instead of *knowing* the parameter values θ_t and θ_b , one *estimates* them from the measurement \mathbf{x} , using maximum likelihood.

Note that whereas the BLR requires the user to specify prior distributions for the unknown parameters, the GLR is a recipe that can be followed without thinking, and it produces an unambiguous result. What we have lost in flexibility and mathematical clarity, we have gained in simplicity.

It is natural to ask whether the GLR is just a special case of the BLR. Schaum^{13,14} provides examples in which the GLR solution can be derived from the BLR formalism, and argues that this provides insight into what the GLR is optimizing. He also describes the ULR (uniform likelihood ratio); this is the special case of the BLR with flat priors: $\pi(\theta)$ constant. But it is not always possible to express the GLR in terms of the BLR; Section 4.2 provides an example where there are no priors that could be used to produce the GLR solution from the BLR formalism.

3.4. Clairvoyant fusion

A generalization of the GLR was proposed by Schaum.¹ To motivate it, I will (following later work by Schaum³⁻⁵ and Bajorski^{9,10}) write the GLR as a min-max optimization over the clairvoyant detectors:

$$\mathcal{T}_{\text{GLR}}(\mathbf{x}) = \frac{\max_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t)}{\max_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b)} = \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \frac{p_t(\mathbf{x}; \theta_t)}{p_b(\mathbf{x}; \theta_b)} = \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}; \theta_t, \theta_b) \leq \lambda. \quad (12)$$

For standard GLR, the statistic $\mathcal{T}_{\text{GLR}}(\mathbf{x})$ is compared to a threshold λ that is fixed. The clairvoyant fusion idea is to make λ depend on the parameters θ_t and θ_b , but this requires that the threshold be brought “inside” the min-max operator:

$$\mathcal{T}_{\text{CF}}(\mathbf{x}) = \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}; \theta_t, \theta_b) / \lambda(\theta_t, \theta_b) \leq 1. \quad (13)$$

If we restrict consideration to functions that are separable in terms of the target and background parameters[†] – *i.e.*, $\lambda(\theta_t, \theta_b) = \lambda_o \lambda_b(\theta_b) / \lambda_t(\theta_t)$ – then we can write

$$p'_t(\mathbf{x}; \theta_t) = p_t(\mathbf{x}; \theta_t) \lambda_t(\theta_t) \quad (14)$$

$$p'_b(\mathbf{x}; \theta_b) = p_b(\mathbf{x}; \theta_b) \lambda_b(\theta_b) \quad (15)$$

and in terms of these scaled (or “penalized”) likelihoods, clairvoyant fusion

$$\mathcal{T}_{\text{CF}}(\mathbf{x}) = \frac{\max_{\theta_t \in \Theta_t} p'_t(\mathbf{x}; \theta_t)}{\max_{\theta_b \in \Theta_b} p'_b(\mathbf{x}; \theta_b)} \leq \lambda_o, \quad (16)$$

looks like the ordinary GLR. This approach also goes by the name *penalized likelihood ratio*.^{15, 16}

If we interpret the scaling factors in terms of priors, $\pi(\theta) \propto \lambda(\theta)$, then

$$\mathcal{T}_{\text{CF}}(\mathbf{x}) = \frac{\max_{\theta_t \in \Theta_t} p_t(\mathbf{x}; \theta_t) \pi_t(\theta_t)}{\max_{\theta_b \in \Theta_b} p_b(\mathbf{x}; \theta_b) \pi_b(\theta_b)} \leq \lambda_o, \quad (17)$$

which looks like BLR, but without the integrals.[‡]

Different strategies for choosing $\lambda(\theta_t, \theta_b)$ are called *flavors*.¹ The (informal) interpretation of those functions in terms of priors on θ_t and θ_b may provide some guidance to that choice. As with Bayesian priors, however, we may be giving the user more flexibility than the user knows what to do with. Just as choosing priors opens a Pandora’s box of possibilities, so also might choosing flavors.

But at least two of the clairvoyant fusion flavors (CF-cfar and CF-cpd) are fully prescriptive. Like the GLR, they do not require any subjective choices to be made by the users. They are in practice a little more complicated to implement than the GLR (although efficient algorithms have been suggested¹⁷), but they provide genuine unambiguous alternatives.

3.4.1. CF-cfar and CF-cpd flavors

One of the clairvoyant fusion flavors that is worth calling out specifically merges detectors with constant false alarm rate. Here $\lambda(\theta_t, \theta_b)$ is chosen so that the associated clairvoyant detector has a given false alarm rate α . Specifically,

$$\lambda(\theta_t, \theta_b, \alpha) = \min_{\lambda} \text{ s.t. } \int_{\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) > \lambda} p_b(\mathbf{x}, \theta_b) d\mathbf{x} \leq \alpha. \quad (18)$$

[†]For example, the L^3R model of Schaum and Daniel^{4,8} is of this form.

[‡]Of course, without the integrals, it’s *not* BLR.

Given this expression for λ , the binary detector takes the form

$$\mathcal{B}_{\text{CF}}(\mathbf{x}, \alpha) = \left\{ \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}(\mathbf{x}, \theta_t, \theta_b) / \lambda(\theta_t, \theta_b, \alpha) \leq 1 \right\}. \quad (19)$$

Schaum has shown how this implicit integral condition can be re-expressed in terms of partial differential equations.¹ And in some cases, analytical expressions can be found.¹ In general, however, Eq. (18) does not permit a simple solution. Further, although the individual detectors have been calibrated to have individual false alarm rates of α , the false alarm rate for the fused detector will be larger than that. To produce a CF detector with a specified false alarm rate, it is generally a matter of trial and error to find the α that leads to this specified rate.

The natural counterpoint to the CF-cfar flavor is one based on P_d instead of P_{fa} . In this case,

$$\lambda(\theta_t, \theta_b, \beta) = \max_{\lambda} \text{ s.t. } \int_{\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) > \lambda} p_t(\mathbf{x}, \theta_t) d\mathbf{x} \geq \beta \quad (20)$$

ensures that the fused decision rules all have a detection rate of β .

3.4.2. Formulation based on monotonic recalibration

Another way to think about the fusion suggested by Eq. (13), for instance, is to observe that just as $\mathcal{T}(\mathbf{x}; \theta_t, \theta_b)$ is a clairvoyant statistic, so is $\mathcal{T}(\mathbf{x}; \theta_t, \theta_b) / \lambda(\theta_t, \theta_b)$; thus Eq. (13) is *still* a min-max operator applied to clairvoyant detectors. But now consider the scalar function $h(z, \theta_t, \theta_b)$ that is monotonic in its first argument: *i.e.*, $z_1 > z_2$ implies $h(z_1, \theta_t, \theta_b) > h(z_2, \theta_t, \theta_b)$. Then it is clear that $\mathcal{T}^*(\mathbf{x}; \theta_t, \theta_b) = h(\mathcal{T}(\mathbf{x}; \theta_t, \theta_b), \theta_t, \theta_b)$ is a clairvoyant statistic. One can then fuse this family of clairvoyant statistic to obtain:

$$\mathcal{T}_{\text{CF}}(\mathbf{x}) = \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} \mathcal{T}^*(\mathbf{x}; \theta_t, \theta_b) \quad (21)$$

$$= \min_{\theta_b \in \Theta_b} \max_{\theta_t \in \Theta_t} h(\mathcal{T}(\mathbf{x}; \theta_t, \theta_b), \theta_t, \theta_b). \quad (22)$$

Although this generalized formulation provides new functions $\mathcal{T}_{\text{CF}}(\mathbf{x})$, and this can in turn be used to develop efficient implementation for CF-cfar and CF-cpd flavors, it does not actually produce any new decision rules.¹⁷

4. ADMISSIBILITY OF HYPOTHESIS TESTING STATISTICS

4.1. Special case of the affine subspace model

Introduced in Ref. [18], the affine subspace model is similar to the additive target model, but with a nonzero offset.

$$\text{H}_0 : \mathbf{x}_b \sim p_b(\mathbf{x}) \quad (23)$$

$$\text{H}_1 : \mathbf{x}_t = \mathbf{x}_b + \mathbf{s}_o + \theta_t \mathbf{s} \sim p_t(\mathbf{x}) = p_b(\mathbf{x} - \mathbf{s}_o - \theta_t \mathbf{s}) \quad (24)$$

Here \mathbf{s}_o and \mathbf{s} are known, but θ_t is unknown. Often there is a bound on permissible values of θ_t ; and by appropriate choice of \mathbf{s}_o and \mathbf{s} , we can rescale those bounds to $-1 \leq \theta_t \leq 1$.

We will consider a special case of the affine subspace model, illustrated in Fig. 1. Here, the composite alternative (target) hypothesis has only two members; specifically, we will take $\theta_t \in \Theta_t = \{\pm 1\}$. We will furthermore consider a two-dimensional space, with $\mathbf{x} = (x, y)^T$, and with $\mathbf{s}_o = (0, 1)^T$ and $\mathbf{s} = (1, 0)^T$. Finally, we will take $p_b(\mathbf{x})$ to be two-dimensional Gaussian centered at the origin, and so $p_t(\mathbf{x}; \theta_t)$ is a Gaussian centered at $(\theta_t, 1)$.

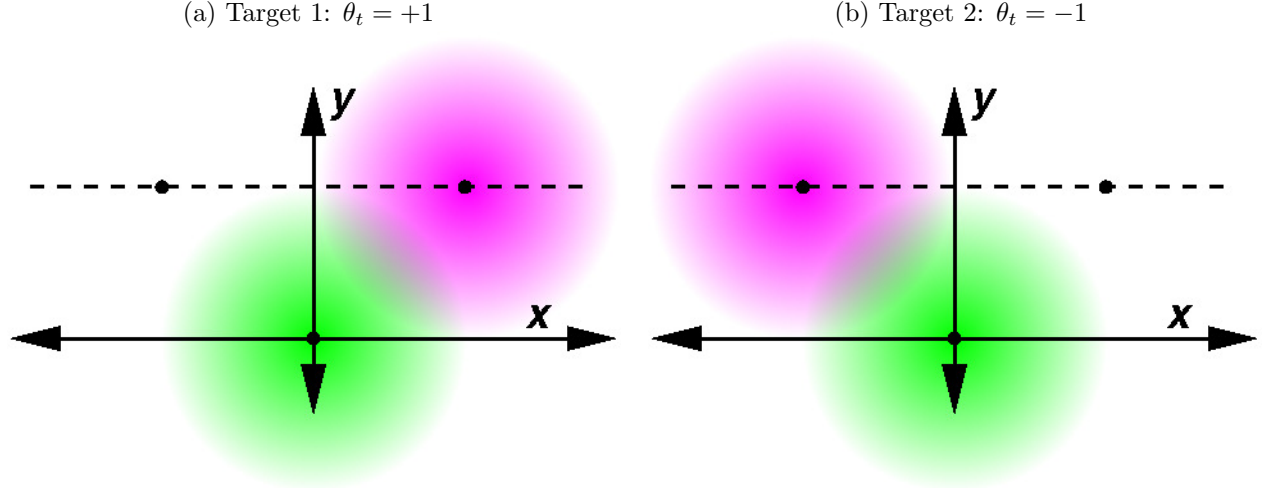


Figure 1. Special case of the affine subspace model. In this model, the composite alternative hypothesis has only two components, shown in (a) and (b). In both cases, the null hypothesis is a unit Gaussian centered at the origin. The alternative hypothesis is a Gaussian centered at $(\theta_t, 1)$ where $\theta_t \in \{-1, +1\}$.

4.2. GLR is not a special case of BLR

For this special case of the affine subspace model, we can explicitly write the two clairvoyant statistics:

$$\mathcal{T}(\mathbf{x}, +1) = \frac{p_t(x, y; +1)}{p_b(x, y)} = \frac{\exp(-((x-1)^2 + (y-1)^2)/2)}{\exp(-(x^2 + y^2)/2)} = \exp(y + x - 1) \quad (25)$$

$$\mathcal{T}(\mathbf{x}, -1) = \frac{p_t(x, y; -1)}{p_b(x, y)} = \frac{\exp(-((x+1)^2 + (y-1)^2)/2)}{\exp(-(x^2 + y^2)/2)} = \exp(y - x - 1) \quad (26)$$

from which it follows that

$$\mathcal{T}_{\text{GLR}}(x, y) = \max_{\theta_t \in \{\pm 1\}} \exp(y + \theta_t x - 1) = \exp(y + |x| - 1). \quad (27)$$

Equivalently, $\mathcal{T}_{\text{GLR}}(x, y) = y + |x|$. Meanwhile, the Bayesian solution is given by

$$\mathcal{T}_{\text{BLR}}(x, y) = \pi_+ \exp(y + x - 1) + \pi_- \exp(y - x - 1) \quad (28)$$

where π_+ and π_- are the prior probabilities assigned to the two targets. Equivalently, $\mathcal{T}_{\text{BLR}}(x, y) = y + \log(\pi_+ \exp(x) + \pi_- \exp(-x))$. Because the alternative hypothesis has only two components, this encompasses *all* Bayesian solutions. It is clear that there is no choice of prior for which π_+, π_- is \mathcal{T}_{BLR} equivalent to \mathcal{T}_{GLR} .

Informally, the difference between \mathcal{T}_{GLR} and \mathcal{T}_{BLR} is the difference between the “maximum” and the “average.” It is also – as seen in Fig. 2(a) – the difference between cuspy contours and smooth contours. Further, as seen in Fig. 2(b), the BLR statistic (at $\pi_+ = \pi_- = 1/2$) is more powerful than the GLR statistic.

4.3. Inadmissibility of clairvoyant fusion

This shows that GLR is not a special case of BLR, but we know that it *is* a special case of CF. More generally, specifying the function $\lambda(\theta_t)$ amounts to choosing two scalars: $\lambda(+1) = \lambda_+$, and $\lambda(-1) = \lambda_-$. In this case,

$$\mathcal{T}_{\text{CF}}(x, y) = y + \log(\max(\lambda_+ \exp(x), \lambda_- \exp(-x))). \quad (29)$$

If we let $x_o = \log(\lambda_-/\lambda_+)/2$, then we can write the equivalent detector

$$\mathcal{T}_{\text{CF}}(x, y) = y + |x - x_o| \quad (30)$$

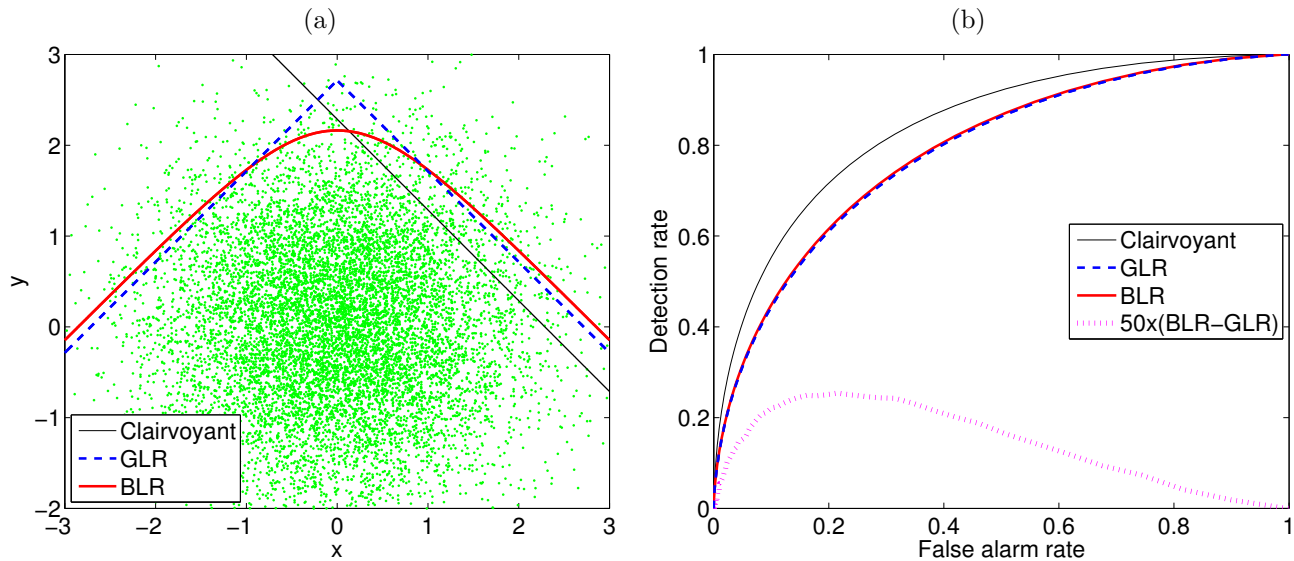


Figure 2. Comparison of BLR and GLR performance for the two-target problem in Section 4.1. (a) In the x - y plane, 10^4 data points are shown for the null hypothesis, along with the three detectors, each calibrated for a false alarm rate of $P_{fa} = 0.05$. The clairvoyant detector has the advantage of knowing which of the two alternative distributions is active. (b) We see that BLR (solid) and GLR (dashed) are both dominated by the clairvoyant detector (thin solid), which is a matched filter that “knows” which of the two targets is active. We also see that although the BLR and GLR have nearly equal performance, the BLR detector is slightly better, and in particular, is better over the whole range of the ROC curve. The dotted curve shows the difference in P_d of BLR minus GLR (what is plotted is $50\times$ that difference); the difference is small but nonnegative. This plot is based on numerical estimates from a sample of a ten million (10^7) points in both the background and the target.

The asymmetry given by $x_o \neq 0$ represents a preference for one of the targets over the other (either because one target is more common than the other, or because detecting one is more valuable than detecting the other). Note, however, that CF-cfar and CF-cpd both retain the symmetry given by $x_o = 0$ and so for this problem are equivalent to the GLR solution.

For this example, neither the set of BLR statistics (parameterized by π_+ and π_-) nor the CF statistics (parameterized by λ_+ and λ_-) are subsets of the other; in fact, the only overlap is at the extremes, which correspond to the two clairvoyant statistics.

Fig. 3 compares the performance of CF and BLR by plotting a measure of performance (P_d at $P_{fa} = 0.05$) for each of the two targets on the two axes. The unambiguous GLR is a single point on the plot, and the CF curve passes through that point. This plot shows, as did Fig. 2(b), that GLR is dominated by a BLR solution; but it also shows that for every CF detector (*i.e.*, for every point on the CF curve), there is a BLR detector (*i.e.*, a point on the BLR curve) that outperforms it. And it outperforms the CF detector *on both targets*.

Thus, for this problem, the GLR detector is not admissible. Further (and this is the bad news), clairvoyant fusion doesn’t fix the problem. Because the alternative hypothesis class only has two members, we can enumerate all of the clairvoyant fusion detectors, and for every one of them, there is a BLR detector that is uniformly more powerful.

This example illustrates two more general principles, described by Lehmann and Romano¹¹ and attributed to Wald.¹⁹ One is that *all* BLR detectors are admissible, and two is that all admissible detectors can be expressed as BLR detectors for some prior (or else can be formed from a limiting process of BLR detectors). If our aim is to optimize the performance of a detector for our particular operational requirements, this says that we can restrict our search to BLR detectors.

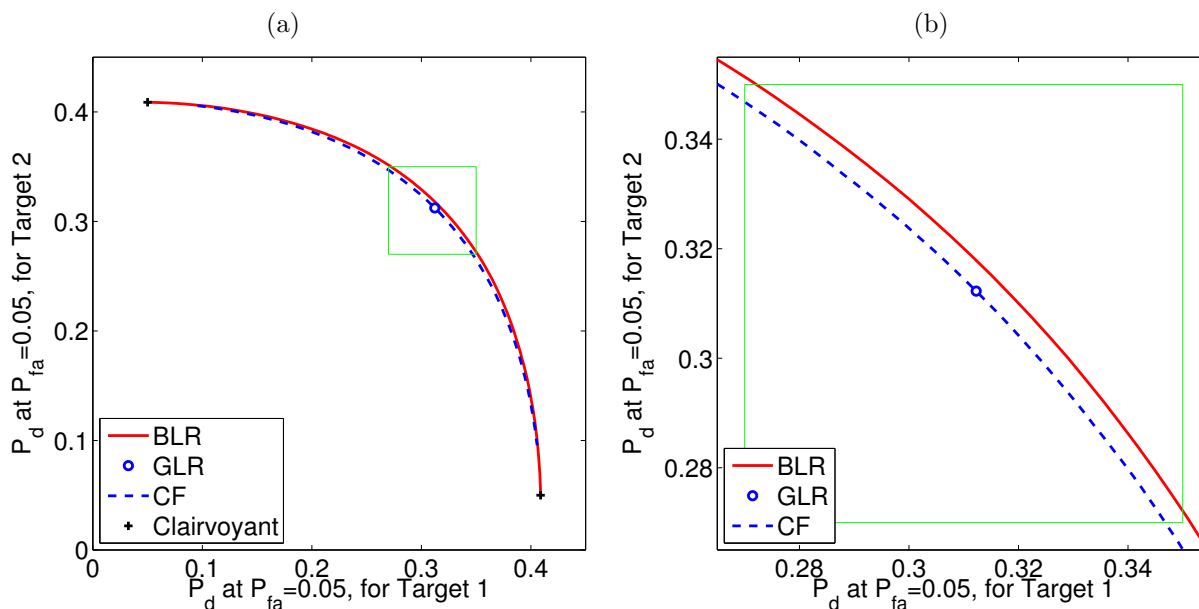


Figure 3. Comparison of BLR and GLR and CF performance for the two-target problem in Section 4.1. Note that both plots show the same data, but (b) is an inset of (a). There are two targets and we do not know which target is present, assuming a target is present at all. The horizontal axis plots the performance against Target 1, and the vertical axis plots the performance against Target 2. As we alter our choice of prior in the BLR, the performance traces out a curve. A prior that is weighted toward Target 1 will get better performance on Target 1, but correspondingly poorer performance on Target 2. The GLR is a single point in this curve and gets equal performance on Target 1 and Target 2, but achieves lower performance than BLR. The CF performance (which includes GLR as a special case) depends on the choice of x_o , Eq. (30), and it traces out a curve whose performance is dominated by the BLR curve. The clairvoyant performance is optimal if one knows which target one is looking for (and dreadful if one guesses incorrectly).

5. CONCLUSIONS

Composite hypothesis testing is an inherently ambiguous (*ergo* confusing) pursuit. By treating this ambiguity as something that can be modeled by probability distributions, the BLR provides a mathematically rigorous solution. But it is a solution that requires the practitioner to put quantitative priors on that subjective ambiguity. As a practical matter, the venerable GLR offers a simpler way to deal with the ambiguity: there is only one GLR detector, and the formulation for producing it is straightforward. But there are no guarantees of optimality for the GLR, and if it produces an unsatisfactory detector, there isn't much one can do about it.

Clairvoyant fusion provides an intriguing alternative that includes GLR as a special case. There are some nice examples where clairvoyant fusion provided a detection that GLR missed,⁸ or a simple formula for a problem in which GLR's solution was intractable.⁶ And in a model that has a UMP solution, the CF-cfar detector retrieved it in its entirety, whereas the GLR detector failed in the $P_{fa} > 0.5$ regime.⁴ These illustrate the potential utility of a method that is not always optimal, and in at least one case, is not even admissible.

What is the future of CF? I am not clairvoyant, but where there is confusion there is curiosity, and I am confident that the composite hypothesis testing problem will continue to engage the research interests of the community. And that better detectors of difficult targets will result from this research.

Acknowledgments

Thanks to Alan Schaum, who introduced me to the idea of continuum fusion, and explained various aspects of CF on a variety of occasions and in multiple venues, from Grenoble and Honolulu to Orlando and Washington DC. I have also benefitted from brief discussions with Peter Bajorski and Brian Daniel. And thanks, too, to Nick Hengartner, who pointed me to the appropriate literature on admissible statistics. This work was supported by the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory.

REFERENCES

1. A. Schaum, "Continuum fusion: A theory of inference, with applications to hyperspectral detection," *Optics Express* **18**, pp. 8171–8181, 2010.
2. A. Schaum, "Continuum fusion: a new methodology for creating hyperspectral detection algorithms," *Proc. SPIE* **7695**, p. 769502, 2010.
3. A. Schaum, "Algorithms with attitude," *Proc. 39th IEEE Applied Imagery Pattern Recognition (AIPR) Workshop*, 2010.
4. A. Schaum, "Continuum fusion: a new methodology for overcoming imperfect sensors and the limitations of statistical detection models," *Proc. MSS (Military Sensing Symposia)*, 2011.
5. A. Schaum, "Design methods for continuum fusion detectors," *Proc. SPIE* **8048**, p. 804803, 2011.
6. A. Schaum, "The continuum fusion theory of signal detection applied to a bi-modal fusion problem," *Proc. SPIE* **8064**, p. 806403, 2011.
7. A. Schaum, "CFAR fusion: A replacement for the generalized likelihood ratio test for Neyman-Pearson problems," *Proc. 40th IEEE Applied Imagery Pattern Recognition (AIPR) Workshop*, 2011.
8. B. Daniel and A. Schaum, "Linear log-likelihood ratio (L^3R) for spectral target detection," *Proc. SPIE* **8048**, p. 804804, 2011.
9. P. Bajorski, "Min-max detection fusion for hyperspectral images," *Proc. 3rd IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2011.
10. P. Bajorski, "Generalized fusion: a new framework for hyperspectral detection," *Proc. SPIE* **8048**, p. 804802, 2011.
11. E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, Springer, New York, 2005.
12. S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. II, Prentice Hall, New Jersey, 1998.
13. A. Schaum, "Spectral subspace matched filtering," *Proc. SPIE* **4381**, pp. 1–17, 2001.
14. A. Schaum, "Hyperspectral target detection using a Bayesian likelihood ratio test," *Proc. IEEE Aerospace Conference* **3**, pp. 1537–1540, 2002.
15. J. Chen, "Penalized likelihood-ratio test for finite mixture models with multinomial observations," *Canadian Journal of Statistics* **26**, pp. 583–599, 1998.
16. A. Vexler, C. Wu, and K. F. Yu, "Optimal hypothesis testing: from semi to fully Bayes factors," *Metrika* **71**, pp. 125–138, 2010.
17. J. Theiler, "Formulation for clairvoyant fusion based on monotonic recalibration of statistics," 2012. Los Alamos National Laboratory Technical Report LA-UR-12-1190.
18. A. Schaum and R. Priest, "The affine matched filter," *Proc. SPIE* **7334**, p. 733403, 2009.
19. A. Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.