ECE 6124
Signal Detection
Introduction

Peter Willett

Fall 2018

1 Basics

1.1 Notional Detection: A Known Additive Signal

Although we will be dealing in discrete time for the most part, let us begin by considering that we have a continuous-time model in which $r(t)$ is the received signal and that we are trying to decide between

$$r(t) \; = \; \nu(t) \tag{1}$$

and

$$r(t) \; = \; s(t) + \nu(t) \tag{2}$$

in which $s(t)$ is a known signal, $\nu(t)$ is a noise process with known statistics, and the observation time is $0 \leq t \leq T_p$. This is vastly over-simplified, since:

- there is no reason to assume that the signal has a known shape;
- even if the signal has a known shape, if may have unknown parameters like amplitude and phase;
- the noise process is in general unknown, for example we may know that it is additive white and Gaussian but may now know its power spectral density;
- many problems do not have a specific starting and ending time of their observation intervals; and
- in any case this " signal plus noise" problem is far too restrictive.

Nonetheless this is a reasonable place to anchor our intuition. And as for the "known-signal" and "specific time" aspects, let us assume that we have an active observation system (radar or sonar) in which a waveform $p(t)$ was

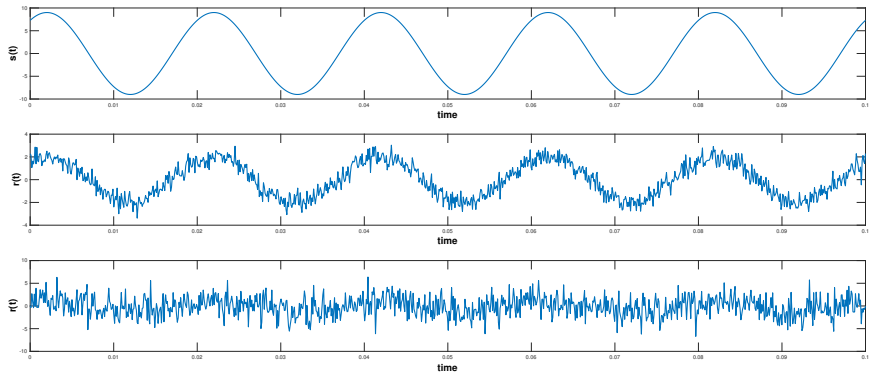

Figure 1: Top row is signal of (4). Middle and bottom rows are corresponding received signals, in cases of moderate and high noise, respectively.

transmitted and $s(t) = p(t - t_0)$, in which t_0 is a suitable time to wait for a return from a target a certain distance away[1].

Suppose you knew that

$$p(t) \;=\; A_t \cos(2\pi f_c t) \tag{3}$$

meaning that

$$s(t) \;=\; A_r \cos(2\pi f_c (t - t_0)) \tag{4}$$

is to be tested for. The top row of figure 1 is what we would expect to see without noise, and the middle and bottom rows are what we might see with moderate and high noise levels, respectively. A fairly obvious strategy would be to form

$$T(r) \;\equiv\; \int_0^{T_p} r(t)s(t)dt \tag{5}$$

and see how big $T(r)$ is. And in fact this turns out to be optimal when $\nu(t)$ is Gaussian: it makes sense to compare what you have to the "template" that you expect to have, and the inner-product (integral) is a good way to perform this comparison.

In fact (5) is optimal regardless of A_r, which is nice. What is less nice is that (5) is in general *not* optimal (or even very good!) when $\nu(t)$ is not Gaussian. And if we got the delay t_0 wrong (5) could be very poor indeed. Consider that we actually receive

$$\begin{aligned} r(t) &= A_r \cos(2\pi f_c (t - t_0')) + \nu(t) &(6)\\ &= A_r \cos(2\pi f_c (t - t_0) + \theta) + \nu(t) &(7) \end{aligned}$$

[1] One may also consider that there be multiple copies of this observation system attuned to different distances: these would be *resolution cells*.

in which $\theta = \pi/2$: the output of (5) is pretty much the same regardless of the presence or absence of signal, since $\cos(2\pi f_c(t-t_0)+\pi/2) = \sin(2\pi f_c(t-t_0))$ and

$$\int_0^{T_p} \cos(2\pi f_c(t-t_0)\sin(2\pi f_c(t-t_0))dt \;\approx\; 0 \tag{8}$$

If $\theta = \pi$ the situation is correspondingly worse.

1.2 A Different Example

Suppose someone has complained to you that dice in a casino are not fair: while actually each face should arise with probability 1/6, the claim is that the dice are weighted such that 3 and 4 have probability 1/4. That would increase the number of 7's that get rolled.

Suppose you roll the pairs of dice $n = 25$ times to test.

Test 1: Since the concern is the number of 7's, count the number of 7's. Report unfairness if that number is too large.

Test 2: Count the number of 3's and 4's, and report unfairness if that number is too large.

Actually these tests both work; but let's see how they do.

We start with Test #2. If the dice are fair, the probability of a 7 is $\frac{1}{6}$ or $\frac{16}{96}$. If they are unfair, that probability rises to $\frac{18}{96}$. The observed number of 7's is distributed binomially in either case. And it's simple to show that if a desired probability of false alarm (that is, of saying the dice are unfair when they are actually fair) is 10%, then the threshold at which unfairness is reported is seven (i.e., if more than seven 7's are observed, you declare that there is cheating going on; otherwise you placate that the game is fair). And if you use this threshold you find that the "probability of detection" – that is, the probability of declaring unfair dice when they are in fact unfair – is about 17%. This isn't good: Test #1 misses most of the cheating.

As for Test #2, we can see that the probability of getting either a 3 or 4 is either $\frac{1}{3}$ or $\frac{1}{2}$. Again using the binomial probability we find the threshold for 10% probability of false alarm is 22 (out of $2n = 50$. And the corresponding probability of detection is about 84%. This is a pretty good test.

So we see that although Test #1 is not completely useless, it really should not be used if Test #2 is available. The reason that Test #1 is not good is that one made an assumption that since 7's were important we should count them. That is, we did not go back to first principles and make clear our model. In doing so we lost valuable information that the observation stream was trying to tell us.

2 Formalization

2.1 Nomenclature

In this course we are interested in deciding between two hypotheses: the *null hypothesis* $\mathcal{H}$ versus the *alternative* $\mathcal{K}$. Sometimes these are denoted $\mathcal{H}_0$ & $\mathcal{H}_1$ – in fact this would make more sense, but since it would require us to carry along a lot of subscripts we will generally work with $\mathcal{H}$ & $\mathcal{K}$. Formally we use $f(x|\theta)$ to denote the probability density (*pdf*) of the observation x, with the understanding that if the observation is discrete $f(\cdot)$ may actually be a probability mass function (*pmf*), and that the observation x can be a scalar, a vector, a time function ... or really anything to which we can assign a probability structure.

We note the appearance of the *parameter*[2] θ. Let us assume that $\theta \in \Theta_H$ or $\theta \in \Theta_K$ – these two sets are exhaustive and mutually exclusive. Then we have the testing problem

$$\begin{aligned} \mathcal{H}: &\quad x \sim f(x|\theta) \quad \theta \in \Theta_H \\ \mathcal{K}: &\quad x \sim f(x|\theta) \quad \theta \in \Theta_K \end{aligned} \tag{9}$$

For example, we might have

$$f(x|\theta) \;=\; \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} \tag{10}$$

and $\Theta_H = \{0\}$ while $\Theta_K = \{1\}$ to indicate a test between $\{x_i\}_{i=1}^n$ being distributed as *iid* unit Gaussian with mean zero versus mean unity. Alternatively, we might have

$$f(x|\theta) \;=\; \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \tag{11}$$

and $\Theta_H = \{\frac{1}{6}\}$ while $\Theta_K = \{\frac{3}{16}\}$ to indicate a test for Bernoulli x_i to be unity with probability $\frac{1}{6}$ versus $\frac{3}{16}$.

Note that both of these examples have both Θ_H and Θ_K singletons. Such tests are called *simple* hypothesis tests, and simple tests are what we will now study. If either one (Θ_H or Θ_K) is a set with more than a single element, the test is *composite*. For example, we could have $\Theta_K = \{x:\ x > 0\}$ in (10) – very easy – or $\Theta_K = \{\frac{1}{16}, \frac{3}{16}\}$ in (11) – more problematic.

[2] Despite the appearance of the vertical bar in $f(x|\theta)$ this is intended only to indicate parametrization.

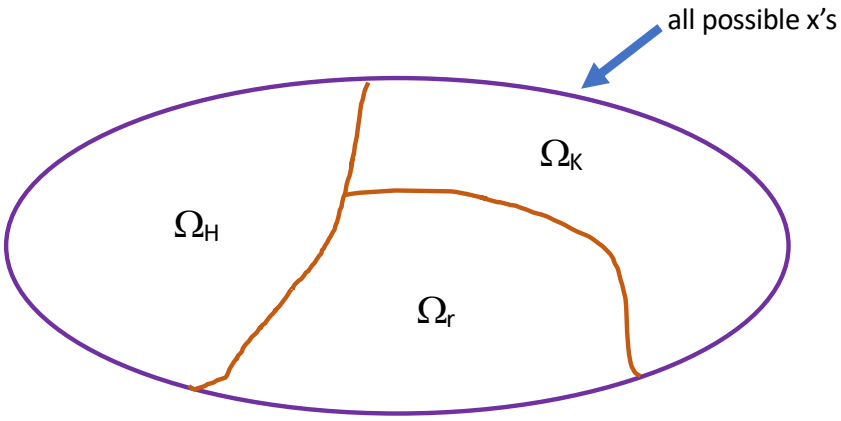

Figure 2: Notional illustration of the three "decision regions" that define the test.

Suppose also that we partition the observation space Ω ($x \in \Omega$) into three exhaustive and disjoint regions Ω_H, Ω_K and Ω_r as in Figure 2. These are actually not necessary for us, but they are intuitive. The real action is that we define the partition function

$$\delta(x) \;=\; \begin{cases} 1 & x \in \Omega_K \\ r(x) & x \in \Omega_r \\ 0 & x \in \Omega_H \end{cases} \tag{12}$$

in which $0 \leq r(x) \leq 1$. Finally we construct the decision $d(x) \in \{\mathcal{H}, \mathcal{K}\}$ as

$$Pr(d(x) = \mathcal{K}) \;=\; \delta(x) \tag{13}$$

That is, if $x \in \Omega_H$ we decide for H and iff $x \in \Omega_K$ we decide for $\mathcal{K}$; iff $x \in \Omega_H$ we "flip a coin" (with probability of a head given by $r(x)$) and decide for $\mathcal{K}$ according to that. This is called randomization, and we will see why it is sometimes necessary.

2.2 What is a Good Test?

2.2.1 Performance to a Statistician

Before we design an optimal test, we need to be clear on what optimal means. One appealing choice is to minimize the probability of error in the decision. Some thought reveals that to do so requires prior probabilities on $\mathcal{H}$ and $\mathcal{K}$; so if we are going this route we may as well also allow for costs of errors

to be different. That is, declaring an emergency (like an imminent airplane crash) when there is none is costly in terms of the emergency response; but it may be far less costly than failing to alert of an emergency situation and thereby failing to respond quickly to it. Such testing is called *Bayesian*, and we will look at it soon.

In many situations no reasonable prior probabilities are available: really, what is the probability of an airplane crash on a certain day? Therefore we will here discuss the *Neyman-Pearson* formulation of detection, in which the goal is to maximize the probability of true detection (i.e., $Pr(d(x) = \mathcal{K}|\mathcal{K})$) subject to a constraint on false-alarm rate ($Pr(d(x) = \mathcal{K}|\mathcal{H}) \leq \alpha_d$).

It is convenient to define the *power function* of the test δ as

$$p(\theta|\delta) \equiv \int \delta(x) f(x|\theta) \tag{14}$$

in which the integral[3] is over all possible observations x. In the simple hypothesis testing situation[4] we have

$$\begin{aligned} p(\theta_H|\delta) &\equiv Pr(d(x) = \mathcal{K}|\mathcal{H}) \\ &= \text{the false alarm rate} \\ &= P_{fa} \\ &= \alpha \\ &= \text{the size of the test} \\ &= \text{the probability of type-I error} \end{aligned} \tag{15}$$

and

$$\begin{aligned} p(\theta_K|\delta) &\equiv Pr(d(x) = \mathcal{K}|\mathcal{K}) \\ &= \text{the detection probability} \\ &= P_d \\ &= \beta \\ &= \text{the power of the test} \\ &= 1 - (\text{the probability of type-II error}) \end{aligned} \tag{16}$$

A good test has large β and small α.

[3]Integration is defined in its most general sense. It could be normal Riemann integration; but it might also a sum, or even a Lebesgue integral.

[4]We note that some (especially statisticians) refer to $p(\theta_K|\delta)$ as $1 - \beta$, versus the definition of β that we will use.

2.2.2 Performance to a Social Scientist

Perhaps this is a making excuses, but probably because they report experimental results that are not really modeled, social scientists prefer not to discuss type-I nor type-II errors. Social scientists generally do have labeled data, however: they know what the true category of each experiment was (sick or well; flora or fauna) and they know what their test selected. At any rate, social scientists use *precision* and *recall*. These are defined as

$$\text{precision} \equiv \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \tag{17}$$

$$\text{recall} \equiv \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \tag{18}$$

These correspond somewhat nicely to what one might see when one is basing one's performance metrics on a *finite* set of experimental outcomes. In these, define

$$n_{HH} = \text{number times } H \text{ declared and } H \text{ true} \tag{19}$$

$$n_{HK} = \text{number times } H \text{ declared and } K \text{ true} \tag{20}$$

$$n_{KH} = \text{number times } K \text{ declared and } H \text{ true} \tag{21}$$

$$n_{KK} = \text{number times } K \text{ declared and } K \text{ true} \tag{22}$$

Then we could equivalently write

$$\text{precision} \equiv \frac{n_{KK}}{n_{KH} + n_{KK}} \tag{23}$$

$$\text{recall} \equiv \frac{n_{KK}}{n_{HK} + n_{KK}} \tag{24}$$

In order to relate these to our statistical viewpoint – which amounts to discarding the idea of a finite sample of experimental outcomes and instead thinking about asymptotic results – we need to define priors: π_H & π_K are the prior probabilities of the two hypotheses. As such, we could write

$$\text{precision} \equiv \frac{\beta\pi_K}{\beta\pi_K + \alpha\pi_H} \tag{25}$$

$$\text{recall} \equiv \frac{\beta\pi_K}{\beta\pi_K + (1-\beta)\pi_H} \tag{26}$$

Personally I find these quantities far more difficult to interpret and use. But there they are.

2.3 The Neyman-Pearson Lemma

The Lemma states that the optimal test, in the sense of maximizing $p(\theta_K|\delta)$ given a constraint on $p(\theta_H|\delta) \leq \alpha_d$ uses

$$\delta(x) \;=\; \left\{ \begin{array}{cc} 1 & f(x|\theta_K) > tf(x|\theta_H) \\ r(x) & f(x|\theta_K) = tf(x|\theta_H) \\ 0 & f(x|\theta_K) < tf(x|\theta_H) \end{array} \right. \tag{27}$$

where the function $r(x)$ and the scalar threshold t are chosen by the constraint.

To show this, write

$$\int \delta(x) f(x|\theta_H) \;=\; \alpha_0 \leq \alpha_d \tag{28}$$

and hence for fixed t

$$\begin{aligned} \beta &= \int \delta(x) f(x|\theta_K) && (29) \\ &= \int \delta(x) f(x|\theta_K) \;+\; t\left[\alpha_0 - \int \delta(x) f(x|\theta_H)\right] && (30) \\ &= t\alpha_0 \;+\; \int \delta(x) \left[f(x|\theta_K) - tf(x|\theta_H)\right] && (31) \end{aligned}$$

The first term is fixed. The second term tells us that δ should be unity when the difference in brackets is positive, and zero when negative. It also tells us that what value δ takes on when the difference in brackets is zero does not matter in the sense that the contribution of the second term to β is null. However, if we can choose $r(x)$ to increase α_0 up to α_d in (28) we increase the term $t\alpha_0$ in (31). The way that $r(x)$ chosen to satisfy this constraint does not matter: one could choose $r(x) = r$ and simple flip a coin; or one could select some region of the set x such that $f(x|\theta_K) = tf(x|\theta_H)$ that will join Ω_K via a binary $r(x)$.

3 The LR and ROC

3.1 The Likelihood Ratio

3.1.1 What It Is

Clearly we have that

$$\{f(x|\theta_K) > tf(x|\theta_H)\} \;\text{ iff }\; \left\{\frac{f(x|\theta_K)}{f(x|\theta_H)} > t\right\} \tag{32}$$

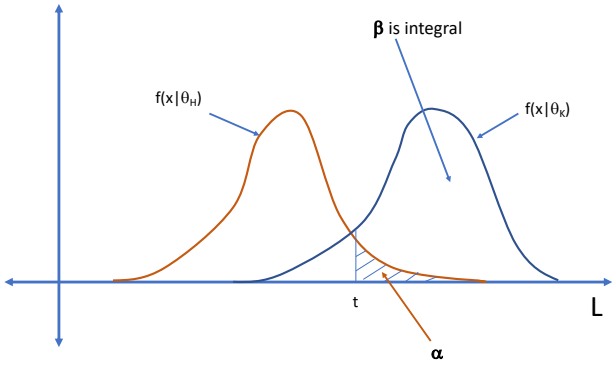

Figure 3: The *pdf*'s of the LR under the two hypotheses. The shaded region indicates the integration of $f(x|\theta_H)$ to get α; the corresponding integral of $f(x|\theta_H K)$ yields β.

where we for convenience define the *likelihood ratio* (LR)

$$L(x) \equiv \frac{f(x|\theta_K)}{f(x|\theta_H)} \tag{33}$$

It is important to recognize that if we have any monotone-increasing function $h(L)$, then we can also write

$$\{L(x) > t\} \text{ iff } \{h(L(x)) > h(t)\} \tag{34}$$

and this may be much simpler. For example, in (10) with $\Theta_H = \{0\}$ while $\Theta_K = \{1\}$ we have

$$L(x) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-1)^2}{2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}} \tag{35}$$

$$= \exp\left(\sum_{i=1}^n x_i - \frac{n}{2}\right) \tag{36}$$

meaning that if we choose

$$h(L) = \frac{1}{n}\log(L) + \frac{1}{2} \tag{37}$$

(which is certainly monotone-increasing) we find the test is simply of the empirical mean versus a threshold. Likewise, in (11) with $\Theta_H = \{\frac{1}{6}\}$ while

$\Theta_K = \{\frac{3}{16}\}$ we have

$$L(x) = \frac{\prod_{i=1}^{n} \left(\frac{3}{16}\right)^{x_i} \left(\frac{13}{16}\right)^{1-x_i}}{\prod_{i=1}^{n} \left(\frac{1}{6}\right)^{x_i} \left(\frac{5}{6}\right)^{1-x_i}} \quad (38)$$

$$= \left(\frac{78}{80}\right)^{n} \left(\frac{15}{13}\right)^{\sum_{i=1}^{n} x_i} \quad (39)$$

meaning that once again with a logarithm and constant we simply need to count up the number of times x_i is unity. Using a logarithm is quite common, and hence we often refer to the *log likelihood ratio* (LLR).

3.1.2 Sufficiency of the Likelihood Ratio

As you know, a statistic that contains all that is needed to be known with respect to estimation of a parameter is call *sufficient*[5]; for example, it can be shown that the empirical average is sufficient for calculation of the mean of a Gaussian. More formally, given an observation x and a parameter ψ, a statistic $T(x)$ is called sufficient for ψ if $f_X(x|T(x) = t)$ does not involve ψ.

A simple way to check for this is via the *factorization theorem* – which is almost just a restatement of the definition – that says that $T(x)$ is sufficient for ψ if and only if

$$f_X(x|\psi) = g(T(x)|\psi)h(x) \quad (40)$$

where g & h are some functions. The "If" part can be see by assuming the factorization can be made and writing

$$f_T(t|\psi) = \int_{x:T(x)=t} f_X(x|\psi) \quad (41)$$

$$= \int_{x:T(x)=t} g(T(x)|\psi)h(x) \quad (42)$$

$$= g(T(x)|\psi) \int_{x:T(x)=t} h(x) \quad (43)$$

which implies

$$f_X(x|T(x) = t, \psi) = \frac{f_{X,T}(x, t|\psi)}{f_T(t, \psi)} \quad (44)$$

$$= \frac{g(T(x)|\psi)h(x)}{g(T(x)|\psi) \int_{x:T(x)=t} h(x)} \quad (45)$$

$$= \frac{h(x)}{\int_{x:T(x)=t} h(x)} \quad (46)$$

[5]Sufficiency means that you might as well throw away the original observations: they give you nothing that you don't already have if you know the sufficient statistic.

where the RHS involves only x. The "Only if" part follows from assuming sufficiency and writing

$$\begin{aligned} f_X(x|\psi) &= fX, T(x,t|\psi) && (47)\\ &= fX(x|t,\psi) f_T(t|\psi) && (48)\\ &= h(x)g(T(x),\psi) && (49) \end{aligned}$$

where the fist line follows from the fact that $T(x)$ is deterministic given x.

For us in this course, the parameter ψ is the hypothesis: $\psi \in \{\mathcal{H}, \mathcal{K}\}$. Now consider that we have

$$\begin{aligned} f_X(x|\psi) &= f_X(x|\mathcal{H}) \begin{cases} L(x) & \psi = \mathcal{K} \\ 1 & \psi = \mathcal{H} \end{cases} && (50)\\ &= h(x)g(L(x),\psi) && (51) \end{aligned}$$

meaning that the likelihood ratio is sufficient for testing. This is, perhaps, not surprising.

3.1.3 Invariance of the Likelihood Ratio

Suppose we write

$$\begin{aligned} f_L(l|\mathcal{K}) &= \int_{x:L(x)=l} f_X(x|\mathcal{K}) && (52)\\ &= \int_{x:L(x)=l} l f_X(x|\mathcal{H}) && (53)\\ &= l \int_{x:L(x)=l} f_X(x|\mathcal{H}) && (54)\\ &= l f_L(l|\mathcal{H}) && (55) \end{aligned}$$

meaning that we can write

$$\frac{f_L(l|\mathcal{K})}{f_L(l|\mathcal{H})} = l \qquad (56)$$

This can be interpreted as meaning that *the likelihood of the likelihood ratio is the likelihood ratio.* This is not surprising given sufficiency, which implies that the LR contains all the relevant information in the data: we should not expect to wring any more information out of it by posing a new hypothesis test based on the LR.

One implication is that we can always write

$$f_L(l|\mathcal{K}) = l f_L(l|\mathcal{H}) \qquad (57)$$

which can be of use when computing performance, since really only one of $\{f_L(l|\mathcal{H}), f_L(l|\mathcal{K})\}$ needs to be computed. And we know that for any *pdf* of L that has mass only on positive values and has mean unity ($\mathcal{E}\{L\} = 1$), this can describe some test via its LR *pdf*.

We also know that an optimal test compares $L(x)$ to a threshold t. We can write

$$\alpha = \int_t^\infty f_L(l|\mathcal{H}) \tag{58}$$

$$\beta = \int_t^\infty f_L(l|\mathcal{K}) \tag{59}$$

We thus have

$$\frac{d\beta}{d\alpha} = \frac{\left(\frac{d\beta}{dt}\right)}{\left(\frac{d\alpha}{dt}\right)} \tag{60}$$

$$= \frac{f_L(t|\mathcal{K})}{f_L(t|\mathcal{H})} \tag{61}$$

$$= t \tag{62}$$

The use of this will be apparent in the next subsection.

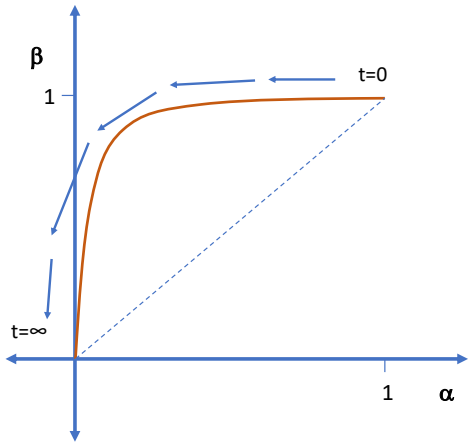

Figure 4: The receiver operating characteristic. The arrows indicate the direction one travels as t increases, see Figure 3.

3.2 The Receiver Operating Characteristic

Now that we know the form of the optimal test, how do we choose this unknown threshold t? In fact, it can be chosen by integration of the LR

pdf as shown in Figure 3. This can, in some cases, be difficult. However, exploring the integrals in Figure 3 and varying t allows us to trace out a performance curve of β versus α, as shown in Figure 4. Note that the upper left is desirable; and that no test can be below the diagonal, since the line $\beta = \alpha$ corresponds to a coin flip that ignores the observation. Figure 4 is known as the *receiver operating characteristic*, more commonly the ROC.

The last likelihood ratio fact from the previous subsection provides us with nice insight: the derivative of the *optimal* ROC is always equal to the threshold used by the (untransformed!) LR. The ROC point furthest to the upper-left – which is graphically-appealing but may not match the false-alarm rate needs – is associated with unity LR threshold ($t = 1$). We also see that if the LR has support over all L the ROC slope begins as ∞ (lower-left) and eventually becomes 0 (upper right). If, for example, the maximum LR possible is 10, then the maximum ROC slope is likewise 10.

3.3 Randomization

The role of randomization – Ω_r, or $\delta(x) = r(x)$ in the Neyman-Pearson Lemma – is probably surprising. Certainly we can see that if the LR contains no point masses (has a density) then there is no reason to consider it. An example where randomization can be ignored is the Gaussian shift-in-mean case (10); and example where it can't is the dice problem (represented by Bernoulli random variables) in (11).

Let us consider three tests such that

$$\delta_1(x) = \begin{cases} 1 & L(x) > t \\ 0 & L(x) \leq t \end{cases} \tag{63}$$

$$\delta_2(x) = \begin{cases} 1 & L(x) > t \\ r(x) & L(x) = t \\ 0 & L(x) < t \end{cases} \tag{64}$$

$$\delta_3(x) = \begin{cases} 1 & L(x) \geq t \\ 0 & L(x) < t \end{cases} \tag{65}$$

such that

$$p(\theta_H|\delta_1) = \alpha_1 < p(\theta_H|\delta_2) = \alpha_2 = \alpha_d < p(\theta_H|\delta_3) = \alpha_3 \tag{66}$$

Note that $r(x)$ has been chosen to satisfy the false-alarm rate requirement exactly, and that $\delta_1(x)$ is too conservative with false-alarms. Clearly the test using $\delta_3(x)$ is unacceptable; let's ignore it from now on. Now we have

$$p(\theta_K|\delta_2) = \int_{x:L(x)=t} r(x) f_X(x|\theta_K) + \int_{x:L(x)>t} f_X(x|\theta_K) \tag{67}$$

$$= \int_{x:L(x)=t} r(x) t f_X(x|\theta_H) \; + \; p(\theta_K|\delta_1) \tag{68}$$

$$= t \int_{x:L(x)=t} r(x) f_X(x|\theta_H) \; + \; p(\theta_K|\delta_1) \tag{69}$$

$$= t(\alpha_d - \alpha_1) \; + \; p(\theta_K|\delta_1) \tag{70}$$

This means not only that $\delta_2(x)$ is better than $\delta_1(x)$, but that the ROC is *linear* in $(\alpha_d - \alpha_1)$. A little thought leads us to see that

$$p(\theta_K|\delta_2) \; = \; p(\theta_K|\delta_1) \; + \; \frac{\alpha_d - \alpha_1}{\alpha_3 - \alpha_1}(p(\theta_K|\delta_3) - p(\theta_K|\delta_1)) \tag{71}$$

which tells us that ROC becomes piece-wise linear, as illustrated in Figure 5. An implication is that an optimal ROC must always be concave: if it were not, it could be made to be via randomization.

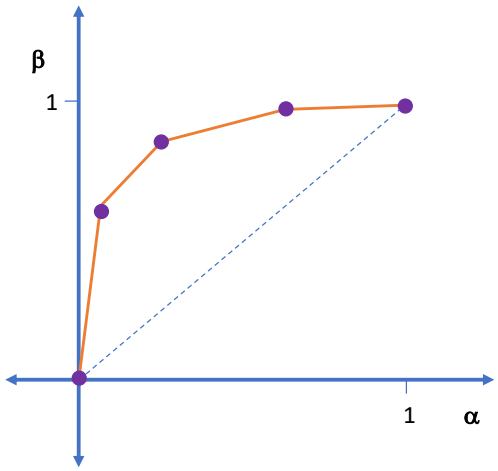

Figure 5: The ROC when there are point-masses. In this case there are five of these. Note that it is possible to have some smooth sections and some piecewise-linear.

4 Examples

4.1 Dice Problem Revisited

Here we reconsider the problem of the dice with too many 7s, caused (perhaps?) by extra weighting for 3's and 4's. The data is the individual rolls $\{x_i\}_{i=1}^{2n}$ where $x_i \in \{1, 2, 3, 4, 5, 6\}$ – it is important not to try to out-think

the mathematics by deciding too early what statistics of the data (like only counting the 7's) we should use. We define

$$z_j(x_i) \equiv \begin{cases} 1 & x_i = j \\ 0 & \text{else} \end{cases} \tag{72}$$

which is isomorphic to x_i. We thus have

$$f(x|\theta) = \prod_{i=1}^{2n} \prod_{j=1}^{6} \theta_j^{z_j(x_i)} \tag{73}$$

and

$$\theta \equiv \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{pmatrix} \quad \text{where} \quad \theta_H = \begin{pmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{pmatrix} \qquad \theta_K = \begin{pmatrix} 1/8 \\ 1/8 \\ 1/4 \\ 1/4 \\ 1/8 \\ 1/8 \end{pmatrix} \tag{74}$$

We write

$$L(x) = \frac{\prod_{i=1}^{2n} \left(\frac{1}{8}\right)^{z_1(x_i)+z_2(x_i)+z_5(x_i)+z_6(x_i)} \left(\frac{1}{4}\right)^{z_3(x_i)+z_4(x_i)}}{\prod_{i=1}^{2n} \left(\frac{1}{6}\right)^{z_1(x_i)+z_2(x_i)+z_3(x_i)+z_4(x_i)+z_5(x_i)+z_6(x_i)}} \tag{75}$$

$$= \prod_{i=1}^{2n} \left(\frac{3}{2}\right)^{z_3(x_i)+z_4(x_i)} \left(\frac{3}{4}\right)^{1-z_3(x_i)-z_4(x_i)} \tag{76}$$

$$= \prod_{i=1}^{2n} \left[\left(\frac{3}{4}\right) (2)^{z_3(x_i)+z_4(x_i)} \right] \tag{77}$$

From this we see that the proper test statistic, is

$$T(x) = \sum_{i=1}^{2n} z_3(x_i) + z_4(x_i) \tag{78}$$

meaning that the observed number of 3's and 4's is to be compared to a threshold.

4.2 Gaussian / Gaussian Problem

$$\begin{aligned} \mathcal{H}: \quad & x_i = \nu_i \\ \mathcal{K}: \quad & x_i = s_i + \nu_i \end{aligned} \tag{79}$$

where the ν_i's are *iid* Gaussian with mean zero and variance σ^2, and the s_i's are independent and Gaussian, the i^{th} with mean μ_i and variance σ_i^2. In our formalism,

$$f(x|\theta) \;=\; \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_{2i}^2}} e^{-(x_i-\theta_{1i})^2/(2\theta_{2i}^2)} \tag{80}$$

and

$$\theta = \begin{pmatrix} \{\mu_i\} \\ \{\sigma_i^2\} \end{pmatrix}: \qquad \Theta_H = \left\{ \begin{pmatrix} \{0\} \\ \{\sigma^2\} \end{pmatrix} \right\} \quad \Theta_K = \left\{ \begin{pmatrix} \{\mu_i\} \\ \{\sigma^2+\sigma_i^2\} \end{pmatrix} \right\} \tag{81}$$

We form the LR and simplify to get

$$T(x) \;=\; \sum_{i=1}^{n} x_i^2 \left(\frac{\sigma_i^2}{\sigma^2(\sigma^2+\sigma_i^2)} \right) \;+\; 2\sum_{i=1}^{n} \mu_i x_i \left(\frac{1}{\sigma^2+\sigma_i^2} \right) \tag{82}$$

Let's see what this means.

- In the case that $\sigma_i^2 = 0$ this means that the mean is actually a deterministic "signal". We get the classic *matched filter*

$$T(x) \;=\; \sum_{o=1}^{n} \mu_i x_i \tag{83}$$

- If $\mu_i = 0$ then we have a test that uses the empirical variance:

$$T(x) \;=\; \sum_{i=1}^{n} x_i^2 \left(\frac{\sigma_i^2}{\sigma^2+\sigma_i^2} \right) \tag{84}$$

  Note that the x_i^2's are weighted by a function that increases with the expected variance under K; but that the weighting is not linear in σ_i^2.

- More generally the optimal test blends matched filter and observed-energy contributions.

4.3 Cauchy Problem

Cauchy noise has a "heavy tail," meaning that the probability of getting a large deviation away from the central point is (much) larger than it would be for Gaussian noise. We write

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \\ \mathcal{K}: &\quad x_i \;=\; \theta \;+\; \nu_i \end{aligned} \tag{85}$$

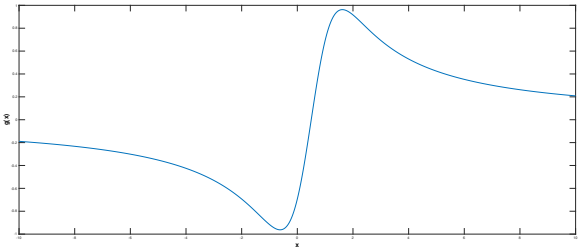

Figure 6: The implied nonlinearity ($g(x)$) that arises when a known unity shift is to be detected in Cauchy noise.

in which the *iid* noise has *pdf*

$$f(\nu) \;=\; \frac{1}{\pi(1+\nu^2)} \tag{86}$$

Formally we have

$$f(x|\theta) \;=\; \prod_{i=1}^{n} \frac{1}{\pi(1+(x_i-\theta)^2)} \tag{87}$$

in which $\Theta_H = \{0\}$ and $\Theta_K = \{1\}$. It is easy to show that the optimal test uses statistic

$$T(x) \;=\; \sum_{i=1}^{n} \log\left(\frac{1+x_i^2}{1+(x_i-1)^2}\right) \;=\; \sum_{i=1}^{n} g(x_i) \tag{88}$$

This is sketched in Figure 6. Note the effect of the heavy-tailed noise: as opposed the Gaussian case, relatively large observations are not especially indicative of the alternative hypothesis. In fact, an observed $x_i = 1$ is far more suggestive of a mean shift than $x_i = 10$.

4.4 Correlated Gaussian Noise

Here we have a known signal $\{s_i\}$ and

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \\ \mathcal{K}: &\quad x_i \;=\; \theta s_i \;+\; \nu_i \end{aligned} \tag{89}$$

in which $\Theta_H = \{0\}$ and $\Theta_K = \{1\}$. The noise is Gaussian, with mean zero and $\mathcal{E}\{\nu_i \nu_j\} = r(i-j)$. It is convenient to write this in terms of vectors as

$$\begin{aligned} \mathcal{H}: &\quad \mathbf{x} \;=\; \nu \\ \mathcal{K}: &\quad \mathbf{x} \;=\; \theta\mathbf{s} \;+\; \nu \end{aligned} \tag{90}$$

where $\mathcal{E}\{\nu\nu^T\} \equiv \mathbf{R}$. Note that $\mathbf{R}$ is Toeplitz. It is straightforward to write the likelihood ratio

$$L(\mathbf{x}) \;=\; \exp\left(\frac{1}{2}\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} - \frac{1}{2}(\mathbf{x}-\mathbf{s})^T\mathbf{R}^{-1}(\mathbf{x}-\mathbf{s})\right) \tag{91}$$

from which we get the test statistic

$$T(\mathbf{x}) \;=\; \mathbf{s}^T\mathbf{R}^{-1}\mathbf{x} \;=\; \mathbf{h}^T\mathbf{x} \tag{92}$$

in which

$$\mathbf{R}\mathbf{h} \;\equiv\; \mathbf{s} \tag{93}$$

We can re-write (92) as

$$T(\mathbf{x}) \;=\; (\mathbf{R}^{-1/2}\mathbf{s})^T(\mathbf{R}^{-1/2}\mathbf{x}) \tag{94}$$

to interpret this matched filter as correlating the *whitened* signal with the whitened observation.

Since $T(\mathbf{x})$ is Gaussian, this is a situation in which we can calculate the performance analytically . . . such cases are, unfortunately, rare. We have

$$\begin{aligned}
\mathcal{E}_H\{T(\mathbf{x})\} &= 0 && (95)\\
\mathcal{V}_H\{T(\mathbf{x})\} &= \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} && (96)\\
\mathcal{E}_K\{T(\mathbf{x})\} &= \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} && (97)\\
\mathcal{V}_K\{T(\mathbf{x})\} &= \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s} && (98)
\end{aligned}$$

in which $\mathcal{V}$ denotes variance. With

$$Q(y) \;\equiv\; \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \tag{99}$$

denoting the unit-Gaussian tail probability, we have

$$\begin{aligned}
\alpha &= Q\left(\frac{\tau}{\sqrt{\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}}\right) && (100)\\
\beta &= Q\left(\frac{\tau - \mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}}\right) && (101)
\end{aligned}$$

or

$$\beta \;=\; Q\left(Q^{-1}(\alpha) - \sqrt{\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}}\right) \tag{102}$$

Note the appearance of the signal-to-noise ratio (SNR) $\mathbf{s}^T\mathbf{R}^{-1}\mathbf{s}$, which we want as large as possible.

If (102) is used for signal design, it is interesting that, ideally, $\mathbf{s}$ turns out to be the eigenvector of $\mathbf{R}$ that corresponds to minimum eigenvalue. If $\mathbf{R}$ is singular the SNR can be infinite; this would imply "perfect detection," which actually makes sense here as signal is designed to lie in a completely noise-free subspace.

It can also easily be shown that the simple white-noise assumption matched filter

$$T(\mathbf{x}) \;=\; \mathbf{s}^T\mathbf{x} \tag{103}$$

has performance

$$\beta \;=\; Q\left(Q^{-1}(\alpha) - \frac{\mathbf{s}^T\mathbf{s}}{\sqrt{\mathbf{s}^T\mathbf{R}\mathbf{s}}}\right) \tag{104}$$

It is interesting that (104) is the same as (102) when $\mathbf{s}$ is an eigenvector of $\mathbf{R}$. (Why is that?)

Also note that the *Toeplitz Distribution Theorem* has it that for Toeplitz matrices of dimension n the eigenvectors converge to the "DFT" sinusoidal vectors, and hence the eigenvalues to the discrete power spectrum samples. That has a nice interpretation for us in signal design: put the signal at the frequency where the noise isn't.

4.5 Gaussian Signal in Gaussian Noise

Here we have

$$\begin{aligned} \mathcal{H}: &\quad \mathbf{x} \;=\; \nu \\ \mathcal{K}: &\quad \mathbf{x} \;=\; \theta\mathbf{s} \;+\; \nu \end{aligned} \tag{105}$$

in which we will take θ as to be determined. The noise ν is Gaussian with covariance $\mathbf{R}_n$ and the signal is likewise Gaussian with covariance $\mathbf{R}_s$; both have mean zero.

We take the likelihood ratio and strip it down to essentials, to get

$$T(\mathbf{x}) \;=\; \mathbf{x}^T\mathbf{R}_n^{-1}\mathbf{x} \;-\; \mathbf{x}^T(\mathbf{R}_n + \theta\mathbf{R}_s)^{-1}\mathbf{x} \tag{106}$$

where of course we could combine these to

$$T(\mathbf{x}) \;=\; \mathbf{x}^T(\mathbf{R}_n^{-1} - (\mathbf{R}_n + \theta\mathbf{R}_s)^{-1})\mathbf{x} \tag{107}$$

for easier computation.

This is interesting, but consider the extreme cases. First we let $\theta \to \infty$, and find

$$T(\mathbf{x}) \;=\; \mathbf{x}^T\mathbf{R}_n^{-1}\mathbf{x} \tag{108}$$

Apparently, in the large-signal case the optimal scheme ignores the structure of the signal entirely and focuses only on seeing whether the observation resembles noise only.

Conversely, let us consider the small-signal case $\theta \to 0$. Let us use the strategy in Golub & Van Loan to help. We know for an invertible matrix $\mathbf{A}(\theta)$ we have

$$\mathbf{A}(\theta)\mathbf{A}(\theta)^{-1} \;=\; \mathbf{I} \tag{109}$$

We differentiate to obtain

$$\frac{d}{d\theta}\left[\mathbf{A}(\theta)\mathbf{A}(\theta)^{-1}\right] \;=\; \mathbf{0} \tag{110}$$

hence

$$\frac{d}{d\theta}\left[\mathbf{A}(\theta)\right]\mathbf{A}(\theta)^{-1} \;+\; \mathbf{A}(\theta)\frac{d}{d\theta}\left[\mathbf{A}(\theta)^{-1}\right] \;=\; \mathbf{0} \tag{111}$$

or

$$\frac{d}{d\theta}\left[\mathbf{A}(\theta)^{-1}\right] \;=\; -\mathbf{A}(\theta)^{-1}\frac{d}{d\theta}\left[\mathbf{A}(\theta)\right]\mathbf{A}(\theta)^{-1} \tag{112}$$

This means that we can write for small θ

$$\frac{d}{d\theta}\left[((\mathbf{R}_n + \theta\mathbf{R}_s)^{-1})\right] \;\approx\; \mathbf{R}_n^{-1} \;-\; \theta\mathbf{R}_s^{-1}\mathbf{R}_s\mathbf{R}_s^{-1} \tag{113}$$

or

$$T(\mathbf{x}) \;=\; \mathbf{x}^T\mathbf{R}_n^{-1}\mathbf{R}_s\mathbf{R}_n^{-1}\mathbf{x} \tag{114}$$

As opposed to the large-signal case in which the optimal strategy is to look for the noise, apparently in the vanishing-signal case the optimal detection structure does pretty close to the opposite: it looks for the (whitened) signal in the (whitened) data.

4.6 Constant Random Signal in White Gaussian Noise

Now we have

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \\ \mathcal{K}: &\quad x_i \;=\; \theta s \;+\; \nu_i \end{aligned} \tag{115}$$

The noise ν_i are *iid* Gaussian with mean zero and unity variance, and the "signal" s is likewise Gaussian with mean zero and unity variance. Note that the signal is unknown but constant.

We have

$$f(\mathbf{x}|s,\theta) \;=\; \left(\frac{1}{2\pi}\right)^n e^{-\sum_{i=1}^n (x_i-\theta s)^2/2} \tag{116}$$

so

$$f(\mathbf{x}|\theta) \;=\; \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{n+1} e^{-\sum_{i=1}^{n}(x_i-\theta s)^2/2 - s^2/2} ds \tag{117}$$

We complete the square to find

$$\begin{aligned}
f(\mathbf{x}|\theta) &= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} e^{-\frac{1}{2}\left[(1+n\theta^2)s^2 - 2(\theta\sum_{i=1}^{n} x_i)s + \sum_{i=1}^{n} x_i^2\right]} ds && (118)\\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} e^{-\frac{(1+n\theta^2)}{2}\left[s^2 - \frac{2}{(1+n\theta^2)}(\theta\sum_{i=1}^{n} x_i)s\right]} \\
&\qquad\qquad \times e^{-\frac{1}{2}(\sum_{i=1}^{n} x_i^2)} ds && (119)\\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} e^{-\frac{(1+n\theta^2)}{2}\left[s^2 - \frac{2}{(1+n\theta^2)}(\theta\sum_{i=1}^{n} x_i)s + \frac{1}{(1+n\theta^2)^2}(\theta\sum_{i=1}^{n} x_i)^2\right]} \\
&\qquad\qquad \times e^{-\frac{1}{2}\left[(\sum_{i=1}^{n} x_i^2) - \frac{1}{(1+n\theta^2)}(\theta\sum_{i=1}^{n} x_i)^2\right]} ds && (120)\\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^{n} \frac{1}{\sqrt{(1+n\theta^2)}} e^{-\frac{1}{2}\left[(\sum_{i=1}^{n} x_i^2) - \frac{1}{(1+n\theta^2)}(\theta\sum_{i=1}^{n} x_i)^2\right]} && (121)
\end{aligned}$$

We take the likelihood ratio and remove superfluous functions, and come up with

$$T(\mathbf{x}) \;=\; \left|\sum_{i=1}^{n} x_i\right| \tag{122}$$

This, if you think about it, makes sense.

5 Bayes Detection

5.1 The Test

Neyman-Pearson detection can be appealing in that very little is needed to specify it beyond the statistical description; in fact, all that is needed is the desired false alarm rate. Bayes detection, on the other hand assumes that much more is available in background. Specifically, assume that one has to hand the prior probabilities of the hypotheses π_H and $\pi_K = 1-\pi_H$. We next assume that we have distributions on $\pi(\theta)$ for all values, such that we could write $\pi(\theta|H)$ and $\pi(\theta|K)$ to describe the conditional behavior of θ under the two hypotheses. These could, of course, be point-masses if θ takes on only one value under each hypothesis.

Further, assume that one has been given *costs* associated with the various decisions: c_{HH}, c_{KH}, c_{HK} & c_{KK}. These should be read such that c_{AB} is the cost of deciding A when in fact B is true. For example, c_{KH} would be

the cost of deciding that K is true when in fact H is true, and might be read as the cost of a *false-alarm*; this might be the cost of sending a fire-truck when there is no fire. Similarly, c_{HK} would be the cost of deciding that H is true when in fact K is true, and might be read as the cost of a *missed detection*; this might be the cost of the burnt building that might have been saved by a fire-truck's dispatch. It is difficult to interpret c_{HH} and c_{KK}, but let us leave them for now.

We define the *loss function*

$$l(\theta, d(x)) \equiv c_{AB} \text{ when } \theta \in \Theta_B \text{ and } d(x) = A \tag{123}$$

We note that all this development could happen nearly equivalently when $l(\theta|d(x))$ varied over θ – this could be reasonable when some values of $\theta \in \Theta_K$ are more costly to miss than others. By extension of (123) we define the *Bayes risk*

$$r(\pi, d) \equiv \mathcal{E}_{x,\theta}\{l(\theta, d(x)\} \tag{124}$$

We expand this latter as

$$\begin{aligned}
r(\pi, d) &= \int_\Theta \int_\Omega l(\theta, d(x)) f(x|\theta)\pi(\theta) dx d\theta \qquad (125)\\
&= \int_{\Theta_H} \left[\int_{\Omega_H} c_{HH} f(x|\theta) dx + \int_{\Omega_K} c_{KH} f(x|\theta) dx \right] \pi(\theta|H)\pi_H d\theta \\
&\quad + \int_{\Theta_K} \left[\int_{\Omega_H} c_{HK} f(x|\theta) dx + \int_{\Omega_K} c_{KK} f(x|\theta) dx \right] \pi(\theta|K)\pi_K d\theta \\
&= c_{HH}\pi_H + \int_{\Theta_H} \left[\int_{\Omega_K} (c_{KH} - c_{HH}) f(x|\theta) dx \right] \pi(\theta|H)\pi_H d\theta \\
&\quad + c_{HK}\pi_K + \int_{\Theta_K} \left[\int_{\Omega_K} (c_{KK} - c_{HK}) f(x|\theta) dx \right] \pi(\theta|K)\pi_K d\theta \\
&= c_{HH}\pi_H + \int_{\Omega_K} \left[(c_{KH} - c_{HH})\pi_H \int_{\Theta_H} f(x|\theta)\pi(\theta|H) d\theta \right] dx \\
&\quad + c_{HK}\pi_K + \int_{\Omega_K} \left[(c_{KK} - c_{HK})\pi_K \int_{\Theta_K} f(x|\theta)\pi(\theta|K) d\theta \right] dx \\
&= c_{HH}\pi_H + \int_{\Omega_K} \left[(c_{KH} - c_{HH})\pi_H f(x|H) \right] dx \\
&\quad + c_{HK}\pi_K + \int_{\Omega_K} \left[(c_{KK} - c_{HK})\pi_K f(x|K) \right] dx \\
&= c_{HH}\pi_H + c_{HK}\pi_K \qquad (126)\\
&\quad + \int_{\Omega_K} \left[(c_{KH} - c_{HH})\pi_H f(x|H) - (c_{HK} - c_{KK})\pi_K f(x|K) \right] dx
\end{aligned}$$

in which

$$f(x|H) \equiv \int_{\Theta_H} f(x|\theta)\pi(\theta|H)d\theta \tag{127}$$

$$f(x|K) \equiv \int_{\Theta_K} f(x|\theta)\pi(\theta|K)d\theta \tag{128}$$

From (126) the optimal strategy to minimize $r(\pi, d)$ is clear. It is

$$d(x) = \begin{cases} K & \frac{f(x|K)}{f(x|H)} \geq \frac{(c_{KH}-c_{HH})\pi_H}{(c_{HK}-c_{KK})\pi_K} \\ H & \frac{f(x|K)}{f(x|H)} < \frac{(c_{KH}-c_{HH})\pi_H}{(c_{HK}-c_{KK})\pi_K} \end{cases} \tag{129}$$

where in fact the risk is not affected the choice what to do at the threshold. Note that this is a *likelihood ratio test*, but differs from the Neyman-Pearson case in that the threshold is fixed, specified and *easy to calculate* from the problem parameters.

As regards this nice threshold

$$\tau = \frac{(c_{KH} - c_{HH})\pi_H}{(c_{HK} - c_{KK})\pi_K} \tag{130}$$

note that if you decide to modify the likelihood ratio $L(x)$ by a monotone increasing transformation to $L(x) \to g(L(x))$, the same transformation must be applied to the threshold $\tau \to g(\tau)$. Note also that these strange costs of making the right decision (c_{HH} & c_{KK}) appear only in an additive way. For example, if you decided that the cost of dispatching a fire-truck to a false alarm was $c_{KH} = \$800$ and (for some unearthly reason) that the cost of not dispatching the fire-truck was $c_{HH} = \$100$, this is operationally the same as $c_{KH} = \$700$ and $c_{HH} = \$0$. As a final note, in the appealing case that the error costs are equal we use

$$\tau = \frac{\pi_H}{\pi_K} \tag{131}$$

in which case the Bayes detector minimizes the probability of error.

5.2 An Example

Suppose we have

$$\begin{aligned} \mathcal{H}: \quad & x_i \sim e^{-x_i}u(x_i) \\ \mathcal{K}: \quad & x_i \sim \frac{1}{2}e^{-x_i/2}u(x_i) \end{aligned} \tag{132}$$

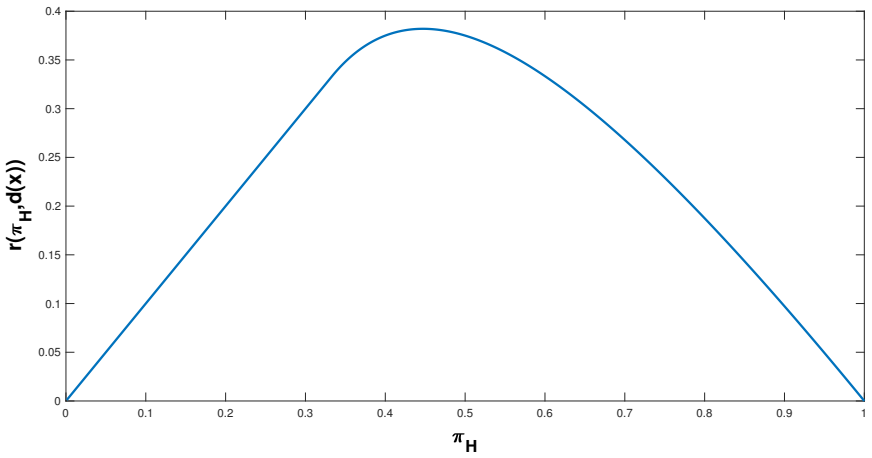

Figure 7: The Bayes risk plotted for the example of (132). The maximum occurs at $\pi_H = 1/\sqrt{5}$.

with independent samples. The costs of errors are equal, here: $c_{KH} = c_{HK}$. The test decides for K if and only if

$$L(\mathbf{x}) \;=\; \left(\frac{1}{2}\right)^n e^{\frac{1}{2}\sum_{i=1}^n x_i} \;\geq\; \frac{\pi_H}{1-\pi_H} \tag{133}$$

or equivalently

$$\sum_{i=1}^n x_i \;\geq\; 2\log\left(2^n \frac{\pi_H}{1-\pi_H}\right) \tag{134}$$

Let us examine this for $n = 1$, in which case we have a threshold test on x:

$$x \;\geq\; 2\log\left(\frac{2\pi_H}{1-\pi_H}\right) \;\equiv\; \tau \tag{135}$$

We first note that if $\pi_H < 1/3$ the RHS is negative: this means that in for such π_H the optimal decision is *always* for $\mathcal{K}$! The corresponding Bayes risk is hence the probability that the hypothesis is $\mathcal{H}$ (since our decision is always for $\mathcal{K}$), meaning $r(\pi, d) = \pi_H$.

When $\pi_H \geq 1/3$ things are more interesting. We have

$$\begin{aligned} r(\pi,d) &= Pr(d(x) = K|H)\pi_H \;+\; Pr(d(x) = H|K)(1-\pi_H) && (136)\\ &= \int_\tau^\infty e^{-x}dx\pi_H \;+\; \int_0^\tau \frac{1}{2}e^{-x/2}dx(1-\pi_H) && (137)\\ &= e^{-2\log(\pi_H/(1-\pi_H))}\pi_H \\ &\qquad + \left(1 - e^{-2\log(\pi_H/(1-\pi_H))/2}\right)(1-\pi_H) && (138) \end{aligned}$$

$$= \left(\frac{1-\pi_H}{2\pi_H}\right)^2 \pi_H \; + \left(\frac{3\pi_H - 1}{2\pi_H}\right)(1-\pi_H) \tag{139}$$

$$= \frac{(1-\pi_H)(5\pi_K - 1)}{4\pi_H} \tag{140}$$

This is plotted in figure 7.

6 Composite Tests

6.1 The Problem

Composite tests are those in which either or both of Ω_H or Ω_K is not a singleton. A benign example is

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \\ \mathcal{K}: &\quad x_i \;=\; \theta \;+\; \nu_i \end{aligned} \tag{141}$$

The noises ν_i are *iid* Laplace, such that

$$f(\mathbf{x}|\theta) \;=\; \prod_{i=1}^{n} \frac{1}{2} e^{-|x-\theta|} \tag{142}$$

for which the optimal test uses the nonlinearity plotted in figure 8. Ignore the vertical extent, since this can be scaled; the important issue is the breakpoint. One might reasonably expect that the test that uses $\theta = 1.7$ when in fact $\theta = 0.9$ would perform suboptimally but not especially poorly. The composite test here to consider might be $\Theta_H = \{0\}$ versus $\Theta_K = \{\theta : \theta > 0\}$.

Conversely, consider the test

$$\begin{aligned} \mathcal{H}: &\quad \mathbf{x} \;=\; \nu \\ \mathcal{K}: &\quad \mathbf{x} \;=\; \mathbf{e}_\theta \;+\; \nu \end{aligned} \tag{143}$$

in which $\mathbf{e}_j$ is the j^{th} Cartesian basis vector[6] and $\Theta_K = \{1, 2, \ldots, n\}$ – we can assume that the noise ν is white and unit Gaussian. It's fairly clear that the test that assumes $\theta = 1$ when in fact $\theta = 2$ would be completely useless.

So, composite tests can either present small issues or be highly pernicious. In the following we shall suggest means to deal with them. Before we do so, please note that both simple and composite tests can be exacerbated by the problem of *nuisance parameters*, these being unknown parameters in the statistical model that do not relate to testing. Consider (143) with

[6]For parallelism, one can take $\Theta_H = \{0\}$ and assign $\mathbf{e}_0$ to be all 0's.

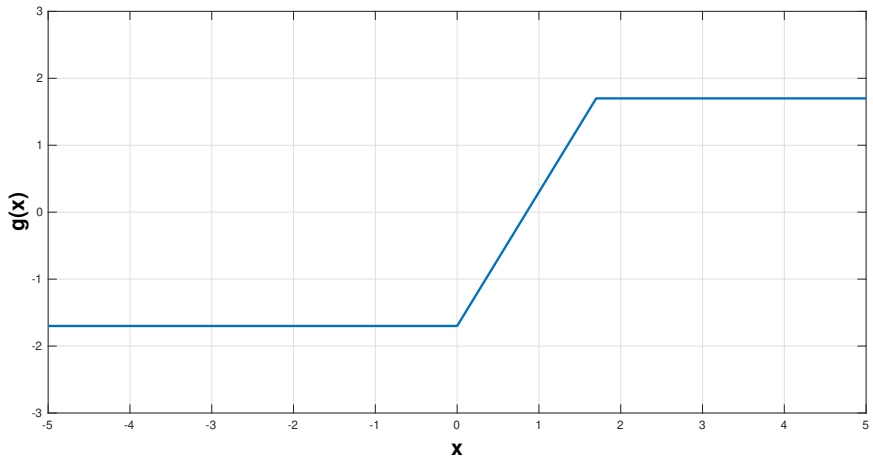

Figure 8: The nonlinearity used for (142). Rather arbitrarily, the value $\theta = 1.7$ has been used here.

$\Theta_K = \{1\}$ (i.e., simple) but with ν of unknown variance σ^2. Unknown parameters cannot be treated using composite hypothesis testing techniques, and we will discuss the matter later in the important case that it relates to the need for testing to be "CFAR" (of constant false-alarm rate).

6.2 Case 1: There is a Prior on θ

To some extent this is tautological, since it is clear that if there is a prior on θ then we can write

$$f(x|H) = \int_{\Theta_H} f(x|\theta) f(\theta|\theta \in \Theta_H) d\theta \tag{144}$$

$$f(x|K) = \int_{\Theta_K} f(x|\theta) f(\theta|\theta \in \Theta_K) d\theta \tag{145}$$

and the hypothesis-testing problem that one might have thought composite reveals itself as simple. An example is suggested by (115), for us here put concretely as

$$\begin{aligned} \mathcal{H}: \quad & x_i = \nu_i \\ \mathcal{K}: \quad & x_i = \theta + \nu_i \end{aligned} \tag{146}$$

where θ is Gaussian. The solution is that the test should use the statistic

$$T(\mathbf{x}) = \left| \sum_{i=1}^{n} x_i \right| \tag{147}$$

as in (122).

6.3 Case 2: A UMP Test Exists

Consider the hypothesis test

$$f(x|\theta) \;=\; \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta s_i)^2}{2}} \tag{148}$$

in which $\{s_i\}$ is a known signal, $\Theta_H = \{0\}$ and $\Theta_K = \{\theta > 0\}$. We select one value of θ, and discover that the optimal LRT can be reduced to comparison of the matched filter

$$T(\mathbf{x}) \;=\; \sum_{i=1}^{n} s_i x_i \tag{149}$$

to a threshold. But we also note that if we had chosen some other $\theta \in \Theta_K$ the form of the test (and the decision) would remain the same. This identifies (148) as being the most powerful (optimal) test for all $\theta \in \Theta_K$: it is *uniformly most powerful*, or UMP. Remember that the likelihood ratio does not define the test; rather $\delta(x)$ – or, if you prefer, Ω_H & Ω_K – defines the test.

UMP tests are wonderful. But they are also, unfortunately, rather rare.

6.4 Case 3: Use a Uniform Prior

If one knows nothing about the relative likelihood of values of $\theta \in \Theta_K$ (or $\theta \in \Theta_H$ for that matter) – meaning one has no reason to expect that any value is more likely than any other value – then it is not illogical to assume that θ has a *uniform* distribution over the set. With this assumed, the hypothesis testing problem reduces precisely to Case 1: this is a simple hypothesis test masquerading as composite.

A nice example of this is the test (143). It is fairly simple to show that the optimal test statistic is

$$T(\mathbf{x}) \;=\; \sum_{i=1}^{n} e^{x_i} \tag{150}$$

which is actually appealing (see later discussion of the GLRT). A problem arises[7] when the set is unbounded, such as in (142) with $\Theta_K = \{\theta > 0\}$. We

[7] Another problem is that there is no universal agreement on what is meant by uniform. For example, some believe that a proper "uniformity" on the variance for a Gaussian random variable ought to imply that the prior distribution of the variance scale as $1/\sigma^2$. Further, mathematically-inclined people become rather tedious about the naturalness of a *conjugate prior*. A conjugate prior remains in the same family when it becomes a posterior. For example, a Gaussian prior with a Gaussian likelihood begets a Gaussian posterior; and a β prior with a Bernoulli likelihood yields a β posterior. Very, very nice and cute and berries & cream to the mathematicians; but why should nature provide these?

would have

$$f(x|K) \;=\; \lim_{A\to\infty} \frac{1}{A} \int_0^A \prod_{i=1}^{n} \frac{1}{2} e^{-|x_i-\theta|} d\theta \tag{151}$$

which converges to zero and has no special useful shape while it does so.

6.5 Case 4: Maximum Likelihood Methods

6.5.1 The GLRT

Here we have, in general, a situation in which neither Θ_H nor Θ_K is a singleton; or at least one of the two is not. Maximum likelihood methods resolve this to simple hypothesis testing by extracting the MLE of θ – separately where appropriate under the two hypotheses – and testing as if the sets were the derived singletons. Mathematically, the generalized likelihood ratio (GLR) is

$$T(x) \;\equiv\; \frac{\max_{\theta\in\Theta_K}\{f(x|\theta)\}}{\max_{\theta\in\Theta_H}\{f(x|\theta)\}} \tag{152}$$

to be compared to a threshold in the usual way. An easy example of the GLR is (143), in which case

$$T(x) \;=\; \max_i\{x_i\} \tag{153}$$

is the GLR. It is interesting to compare (153) to (150); clearly they are different[8] but also just as clearly they focus on the large-value observation. They are especially different in the small signal regime in which (150) is more forgiving.

Another and rather canonical example of the GLR is the *linear model*

$$\mathbf{x} \;=\; \mathbf{A}\theta \,+\, \nu \tag{154}$$

in which ν is white and Gaussian and $\mathbf{A}$ is a known matrix. This can be formalized as

$$f(\mathbf{x}|\theta) \;=\; \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{A}\theta)^T(\mathbf{x}-\mathbf{A}\theta)} \tag{155}$$

with θ an m-vector and where $\Theta_H = \{\mathbf{0}\}$ and $\Theta_K = \{\theta : \theta \neq \mathbf{0}\}$. Provided $m < n$ (n is the length of $\mathbf{x}$) we have the MLE

$$\hat{\theta} \;=\; (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{156}$$

[8] Do not be fooled by the value itself: either exponentiate (153) or take the logarithm of (150) to get them on the same scale, and note that as usual monotone-increasing nonlinearities have no effect on the test.

which means that the "signal" in the derived simple hypothesis test is

$$\hat{s} \;=\; \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{157}$$

and hence the GLR (the derived matched filter) is

$$T(\mathbf{x}) \;=\; \mathbf{x}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{158}$$

which is attractive. We will in a later chapter spend more time with various GLRTs.

The GLR possesses some optimality properties when the situation is asymptotic. But from experience, the GLRT is seldom beaten by other tests in a composite testing situations. Many great statisticians (such as Lehmann) do not even bother with the monicker "generalized" ... they simply call it the likelihood ratio test. So the (rather uninteresting) advice that must be given is that if one can easily derive the GLR for one's problem – this means either that the MLE is explicit or its numerical invocation is well-behaved enough that it is satisfactory to say that $\hat{\theta}$ is in the neighborhood of the true θ – then one probably ought to use the GLRT. It is the other situations, usually small-signal ones where the MLE performance is not so good, that are perhaps more interesting.

As a final note, the situation of testing with nuisance parameters (like the variance) can be posed as a GLR. It does not always work, but can in some cases offer satisfying performance.

6.5.2 Asymptotic GLRs

There are two forms of GLR that are worth mentioning, since they can be simpler to implement than the raw GLR and since they offer some performance expressions. These are both based on the asymptotic properties of the MLE. These are the Wald and Rao tests, and we will discuss them later.

6.6 Case 5: The Small-Signal Approximation

Let us assume that $\Theta_H = \{0\}$ and $\Theta_K = \{\theta : \theta > 0\}$. The GLR approach would estimate θ via maximum likelihood, but if θ is very small such estimation may not work effectively ... or at least it may result in a complicated test.

The *locally-optimal* (LO) approach is to assume that θ is very small, although not 0, under $\mathcal{K}$. The reason is two-fold. First, the test statistic so formed is often very simple and in many cases has good performance predictability. The second reason is more intuitive than provable: a test

that works as well as it can in the most difficult situation (of vanishing signal) *probably* works well (admittedly no longer optimally) when the signal is larger. We will spend a great deal of the course with LO detection.

The key property that informs LO detection is the Taylor series:

$$f(x|\theta) \;\approx\; f(x|\theta)|_{\theta=0} \;+\; \theta \left.\frac{d}{d\theta} f(x|\theta)\right|_{\theta=0} \tag{159}$$

meaning that we can use

$$T(x) \;=\; \left.\frac{\dot{f}(x|\theta)}{f(x|\theta)}\right|_{\theta=0} \tag{160}$$

as our LO test statistic. More on this next chapter.

ECE 6124
Signal Detection
Locally Optimal Detection and Efficacy

Peter Willett

Fall 2018

1 Generalized Neyman-Pearson Lemma

Suppose we want to design a test to maximize

$$\int g(x)\delta(x) \tag{1}$$

such that

$$\int h_1(x)\delta(x) \leq \alpha_1 \tag{2}$$

$$\int h_2(x)\delta(x) \leq \alpha_2 \tag{3}$$

$$\vdots \quad \vdots \quad \vdots \tag{4}$$

Then that test uses

$$\delta(x) = \begin{cases} 1 & g(x) > \sum_i t_i h_i(x) \\ r(x) & g(x) = \sum_i t_i h_i(x) \\ 0 & g(x) < \sum_i t_i h_i(x) \end{cases} \tag{5}$$

for some $\{t_i\}$. The proof is so similar to that of the Neyman-Pearson Lemma that it is not given.

2 Small Signals

2.1 Asymptotic Power Function

Suppose we have the usual hypothesis test

$$x \sim f(x|\theta) \tag{6}$$

where

$$\Theta_H = \{\theta = 0\} \quad \Theta_K = \{\theta : \theta > 0\} \tag{7}$$

Consider the situation of figure 2. This shows the power functions

$$p(\theta|\delta) \equiv \int \delta(x) f(x|\theta) \tag{8}$$

of three tests. Our claim in this section is that all tests are good when θ is large under $\mathcal{K}$; hence we should concentrate on the situation that θ is small. In that regime it appears that δ_1 is best. If indeed it is the best possible test over some (open) interval $0 < \theta < \epsilon$ then we call it *locally-optimal*. That we be able to write

$$p(\theta|\delta) \approx p(0|\delta) + \theta \dot{p}(0|\delta) \tag{9}$$

for small θ, in which

$$\dot{p}(0|\delta) = \frac{d}{d\theta} p(\theta|\delta)\Big|_{\theta=0} \tag{10}$$

is key to our analysis. Clearly our goal is to maximize $\dot{p}(0|\delta)$.

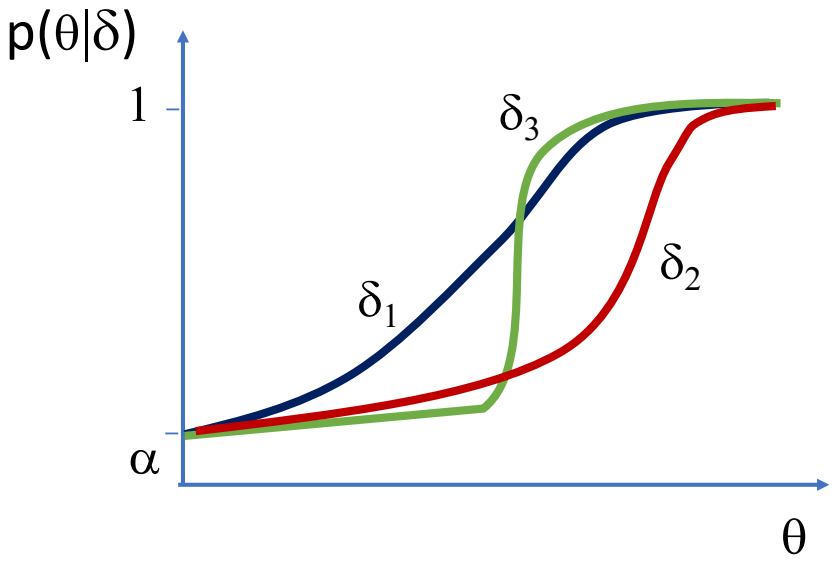

Figure 1: Power functions of three tests.

2.2 One-Sided Alternative: Locally-Optimal Detection

Testing $\Theta_H = \{0\}$ versus $\Theta_K = \{\theta > 0\}$, we desire to maximize

$$\begin{aligned} \dot{p}(0|\delta) &= \frac{d}{d\theta} \int \delta(x) f(x|\theta) \Big|_{\theta=0} && (11) \\ &= \int \delta(x) \frac{d}{d\theta} f(x|\theta) \Big|_{\theta=0} && (12) \\ &\equiv \int \delta(x) \dot{f}(x|0) && (13) \\ &\equiv \int \delta(x) g(x) && (14) \end{aligned}$$

and

$$h(x) = f(x|0) \quad (15)$$

so the *locally-optimal* test uses the thresholded statistic

$$T_{lo}(x) \equiv \frac{\dot{f}(x|0)}{f(x|0)} \quad (16)$$

We have seen this form of tests statistic before, developed as linear terms in the Taylor expansion of the likelihood ratio.

2.3 Two-Sided Alternative: Unbiased Testing

Suppose have $\Theta_H = \{0\}$ and $\Theta_K = \{\theta \neq 0\}$. If we use the LOD formulation to maximize $\dot{p}(\theta|\delta)$ it is mathematically obvious that while this is very good for $\theta > 0$ it is not at all good for $\theta < 0$: when $\theta < 0$

$$\begin{aligned} p(\theta|\delta) &= p(0|\delta) - |\theta| \dot{p}(0|\delta) && (17) \\ &< \dot{p}(0|\delta) && (18) \end{aligned}$$

The condition of *unbiased*ness[1] insists that, at least in the small-signal regime, β is never less than α. Assuming continuity and differentiability, this implies that we must insist that $\dot{p}(0|delta) = 0$. Since for small signals we can write

$$p(\theta|\delta) \approx p(0|\delta) + \theta \dot{p}(0|\delta) + \frac{1}{2}\theta^2 \ddot{p}(0|\delta) \quad (19)$$

this means our goal is

$$\begin{aligned} \text{maximize} \quad & \ddot{p}(0|\delta) && (20) \\ \text{subject to} \quad & \dot{p}(0|\delta) = 0 && (21) \\ \text{and} \quad & p(0|\delta) = \alpha && (22) \end{aligned}$$

[1] The "bias" term is so over-used that I think its reappearance here is a real shame.

The generalized Neyman Pearson Lemma delivers the result: compare

$$T(x) \;=\; \frac{\ddot{f}(x)}{f(x)} \;+\; \lambda \frac{\dot{f}(x)}{f(x)} \tag{23}$$

to a threshold.

2.4 Example of Unbiassed Testing

Suppose we have (the usual, for now) situation that

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \theta s_i \;+\; \nu_i \end{aligned} \tag{24}$$

where the noise $\{\nu_i\}$ are *iid* Gaussian and the signal $\{s_i\}$ is known. The sets are $\Theta_H = \{0\}$ and $\Theta_K = \{\theta \neq 0\}$. We have

$$f(x|\theta) \;=\; \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta s_i)^2\right) \tag{25}$$

from which

$$\dot{f}(x|\theta) \;=\; \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} s_i(x_i - \theta s_i)\right) f(x|\theta) \tag{26}$$

and hence

$$\ddot{f}(x|\theta) \;=\; \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} s_i(x_i - \theta s_i)\right)^2 f(x|\theta) \;-\; \left(\frac{1}{\sigma^2}\sum_{i=1}^{n} s_i^2\right) f(x|\theta) \tag{27}$$

Setting θ to zero and discarding irrelevant terms, we have

$$T_{unb}(x) \;=\; \left(\sum_{i=1}^{n} s_i x_i\right)^2 \;+\; \lambda \left(\sum_{i=1}^{n} s_i x_i\right) \tag{28}$$

Some easy analysis would show that to preserve unbiasedness we must have $\lambda = 0$. We are left with

$$T_{unb}(x) \;=\; \left|\sum_{i=1}^{n} s_i x_i\right| \tag{29}$$

Interestingly, this is the same form that we derived last section for the case that θ was Gaussian.

3 Locally Optimal Detection of Known Signals

3.1 The Nonlinear Correlator

With $\{\nu_i\}$ independent having densities f_i and $\{s_i\}$ known, consider

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \theta s_i \;+\; \nu_i \end{aligned} \tag{30}$$

and $\Theta_H = \{0\}$ versus $\Theta_K = \{\theta > 0\}$ we have

$$f(x|\theta) \;=\; \prod_{i=1}^{n} f_i(x_i - \theta s_i) \tag{31}$$

We take the derivative (product rule) and get

$$\dot{f}(x|\theta) \;=\; \sum_{i=1}^{n} -s_i \dot{f}_i(x_i - \theta s_i) \prod_{j \neq i} f_j(x_j - \theta s_j) \tag{32}$$

which means we use

$$T_{lo}(x) \;=\; \sum_{i=1}^{n} s_i \frac{-\dot{f}_i(x_i)}{f_i(x_i)} \tag{33}$$

In the *iid* case we simply[2] use f and write

$$T_{lo}(x) \;=\; \sum_{i=1}^{n} s_i g_{lo}(x_i) \tag{34}$$

with

$$g_{lo}(x) \;\equiv\; -\frac{\dot{f}(x)}{f(x)} \tag{35}$$

as our test statistic. That is, we "correlate" (i.e., the dot-product) the known signal with a version of the observations that may have been nonlinearly distorted by this function g_{lo}.

3.2 The Gaussian Case

Not surprisingly, with

$$f(x_i) \;=\; \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2} x_i^2} \tag{36}$$

[2]It is admitted that we use f to represent the full pdf of the entire observation stream and the marginal. This is hereby apologized for but not avoided.

we have

$$\dot{f}(x_i) \;=\; -\frac{x_i}{\sigma_i^2} f(x_i) \tag{37}$$

and

$$T_{lo}(x) \;=\; \sum_{i=1}^{n} \frac{s_i x_i}{\sigma_i^2} \tag{38}$$

We have

$$T_{lo}(x) \;=\; \sum_{i=1}^{n} s_i x_i \tag{39}$$

if $\sigma_i = \sigma$ – the *iid* case.

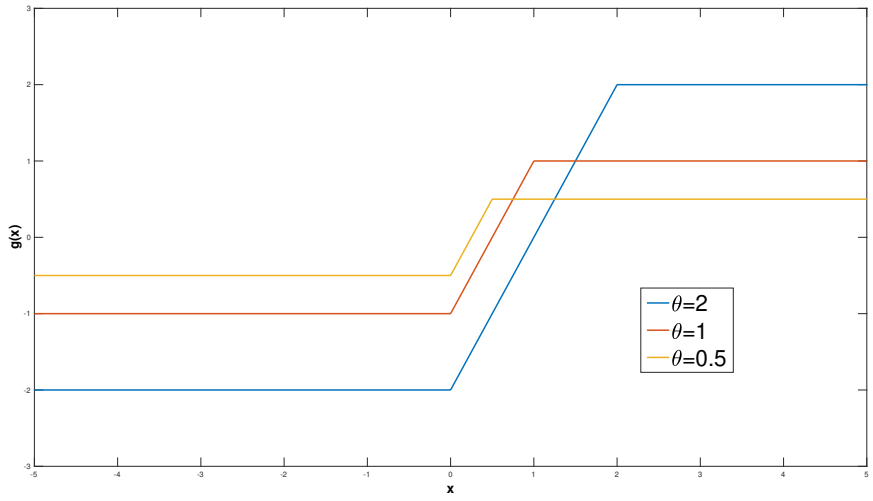

Figure 2: The Laplace nonlinearity for decreasing values of θ and corresponding to $s = 1$, Note the convergence to a signum function. Ignore the vertical scaling: this would be absorbed into the threshold.

3.3 The Laplace Case

Here we have

$$f(x_i) \;=\; \frac{1}{2a} e^{-a|x_i|} \tag{40}$$

from which

$$g_{lo}(x) \;=\; \mathrm{sgn}(x) \tag{41}$$

is easily calculated. See figure 2.

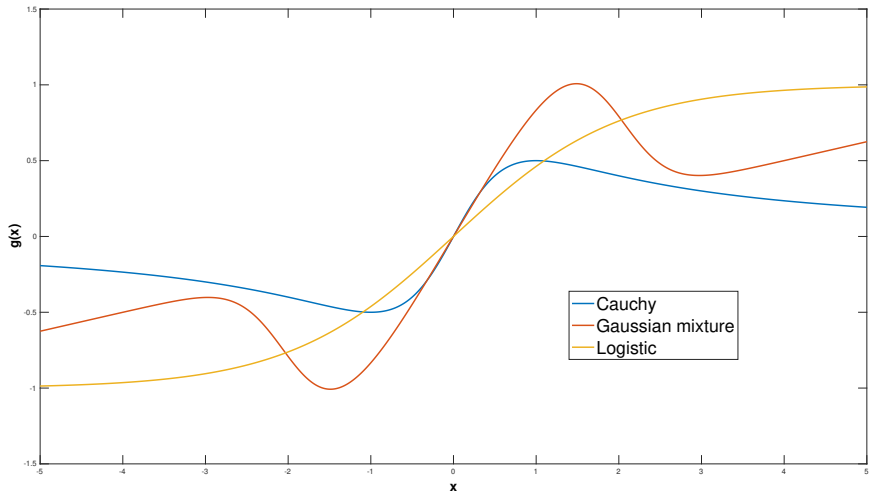

Figure 3: Nonlinearities for Cauchy, Gaussian mixture and Logistic cases. The GM case uses $\varepsilon = 0.1$, $\sigma_1 = 1$ and $\sigma_2 = 2$ – in cases of interest one would be far more likely to find a large ratio σ_2/σ_1, but these are chosen for plotting purposes. Note that the Cauchy g_{lo} is strict about "blanking" large-amplitude observations. The Gaussian mixture g_{lo} begins to rise again and eventually becomes infinite. The logistic can be seen to be a softer version of the Laplace (sign) case.

3.4 The Cauchy Case

Here we have

$$f(x_i) \;=\; \frac{1}{\pi(1+x_i^2)} \tag{42}$$

from which

$$g_{lo}(x) \;=\; \frac{x_i}{(1+x_i^2)} \tag{43}$$

is easily calculated. See figure 3, which compares this to Gaussian mixture and logistic.

3.5 The Gaussian Mixture Case

Here we have

$$f(x_i) \;=\; (1-\varepsilon)\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2\sigma_1^2}x_i^2} \;+\; \varepsilon\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}x_i^2} \tag{44}$$

where typically σ_1 refers to ambient ("well-behaved") noise samples and $\sigma_2 \gg \sigma_1$ refers to outliers – impulsive interference events – that happen

only rarely (a fraction ε of the time). We have

$$g_{lo}(x) \;=\; \frac{(1-\varepsilon)\frac{x}{\sigma_1^3}e^{-\frac{1}{2\sigma_1^2}x^2} \;+\; \varepsilon\frac{x}{\sigma_2^3}e^{-\frac{1}{2\sigma_2^2}x^2}}{(1-\varepsilon)\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2\sigma_1^2}x^2} \;+\; \varepsilon\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{1}{2\sigma_2^2}x^2}} \tag{45}$$

which is probably more complicated than we hoped. See figure 3.

3.6 Logistic Case

Here we have the (familiar) cumulative distribution function

$$F(x) \;=\; \frac{1}{1+e^{-x}} \tag{46}$$

and the probably-less familiar *pdf*

$$f(x) \;=\; \frac{e^{-x}}{(1+e^{-x})^2} \tag{47}$$

We write

$$f(x) \;=\; \frac{1}{1+e^{-x}} \;-\; \frac{1}{(1+e^{-x})^2} \tag{48}$$

and hence

$$\dot{f}(x) \;=\; \frac{e^{-x_i}}{(1+e^{-x})^2} \;-\; \frac{2e^{-x}}{(1+e^{-x})^3} \tag{49}$$

From this we get

$$g_{lo}(x) \;=\; \frac{2}{1+e^{-x}} \;-\; 1 \tag{50}$$

or

$$g_{lo}(x) \;=\; \frac{1-e^{-x}}{1+e^{-x}} \;=\; \tanh(x/2) \tag{51}$$

See figure 3.

4 Comparison of Detectors

4.1 Relative Efficiency anf Asymptotic Relative Efficiency

Suppose we have a hypothesis test, and two different detection schemes, δ_a & δ_b. Both operate at ROC point (α, β), and to do so they require respectively n_a & n_b samples. The the *relative efficiency* of scheme δ_a to δ_b is

$$\mathrm{RE}_{a,b}(\alpha,\beta) \;\equiv\; \frac{n_b}{n_a} \tag{52}$$

Clearly, then, if δ_a requires fewer samples than δ_b, then the relative efficiency is greater than unity, which makes sense. This is fine to write, but RE's utility is hampered by its parametrization in terms of (α, β). The more useful concept follows, and it assumes that there is no such dependence.

To define the *asymptotic relative efficiency* (ARE) we need first to require that $\Theta_H = \{\theta_0\}$ and $\Theta_K = \{\theta_l\}$ where

$$\lim_{l \to \infty} \theta_l \;=\; \theta_0 \tag{53}$$

Unless the situation is trivial, n_a & n_b both diverge as the hypotheses approach one another. Then we have

$$\mathrm{ARE}_{a,b} \;\equiv\; \lim_{l \to \infty} \left\{ \frac{n_b}{n_a} \right\} \tag{54}$$

provided that limit is independent of α & β. It will be seen that some care must be taken with the sequence $\{\theta_l\}$ since if if converges too quickly the only possible value of (α, β) is $\alpha = \beta$ (detection is impossible); and if too slowly, then (almost) any detector can achieve $(\alpha, \beta) = (0, 1)$. That is, care must be taken to avoid triviality; for now (and this will change, later) picture that $\theta_l = \theta_0 + \gamma/\sqrt{l}$, for some γ that does not matter.

For example, let us examine

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \theta \;+\; \nu_i \end{aligned} \tag{55}$$

where the noise $\{nu_i\}$ are *iid* Gaussian $\mathcal{N}(0, \sigma^2)$ and the signal $\{s_i\}$ is known. The sets are $\Theta_H = \{0\}$ and $\Theta_K = \{\theta \neq 0\}$. Detector δ_a is the linear detector, using

$$T_a(x) \;=\; \sum_{i=1}^{n_a} x_i \tag{56}$$

and detector δ_b is the sign detector, with

$$T_b(x) \;=\; \sum_{i=1}^{n_b} \mathrm{sgn}(x_i) \tag{57}$$

Clearly $T_a(x)$ is exactly Gaussian, with

$$\mathcal{E}\{T_a|\mathcal{H}\} \;=\; 0 \tag{58}$$

$$\mathcal{E}\{T_a|\mathcal{K}\} \;=\; n_a\theta \tag{59}$$

$$\mathcal{V}\{T_a|\mathcal{H}\} \;=\; n_a\sigma^2 \tag{60}$$

$$\mathcal{V}\{T_a|\mathcal{K}\} \;=\; n_a\sigma^2 \tag{61}$$

so

$$\beta \;=\; Q\left(Q^{-1}(\alpha) - \frac{n\theta}{\sqrt{n\sigma^2}}\right) \;=\; Q\left(Q^{-1}(\alpha) - \frac{\sqrt{n_a}\theta}{\sigma}\right) \tag{62}$$

as we've seen previously, and Q is, as before, the unit-Gaussian tail probability. As for δ_b, we appeal to the central limit theorem (CLT) and assume $T_b(x)$ is also Gaussian, with

$$\begin{aligned}
\mathcal{E}\{T_b|\mathcal{H}\} &= 0 && (63)\\
\mathcal{E}\{T_b|\mathcal{K}\} &= n_b(2q-1) && (64)\\
\mathcal{V}\{T_b|\mathcal{H}\} &= n_b && (65)\\
\mathcal{V}\{T_b|\mathcal{K}\} &= n_b 4q(1-q) && (66)
\end{aligned}$$

in which

$$\begin{aligned}
q &\equiv Pr(x_i > 0|\mathcal{K}) && (67)\\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\theta)^2/2\sigma^2} dy && (68)\\
&= \frac{1}{2} + \int_{-\theta}^0 \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} dy && (69)\\
&\approx \frac{1}{2} + \theta\frac{1}{\sqrt{2\pi}\sigma} && (70)
\end{aligned}$$

where the approximation assumes that θ is small. Now assuming the CLT we write

$$\alpha \;=\; Q\left(\frac{\tau}{\sqrt{n}})\right) \tag{71}$$

in terms of the necessary threshold τ, so

$$\tau \;=\; \sqrt{n}Q^{-1}(\alpha) \tag{72}$$

hence

$$\beta \;\approx\; Q\left(\frac{\sqrt{n_b}Q^{-1}(\alpha) - n_b(2q-1)}{\sqrt{n4q(1-q)}}\right) \tag{73}$$

Now since from (70) we have

$$\lim_{\theta\to 0}\{4q(1-q)\} \;=\; 1 \tag{74}$$

we can write

$$\beta \;\approx\; Q\left(Q^{-1}(\alpha) - \frac{n_b(2q-1)}{\sqrt{n_b}}\right) \;\approx\; Q\left(Q^{-1}(\alpha) - \sqrt{n_b}\theta\frac{2}{\sqrt{2\pi}\sigma}\right) \tag{75}$$

We equate (62) and (75) and find

$$\frac{\sqrt{n_a}\theta}{\sigma} = \sqrt{n_b}\theta\frac{2}{\sqrt{2\pi}\sigma} \tag{76}$$

or

$$\begin{aligned} \mathrm{ARE}_{a,b} &= \lim_{l\to\infty}\left\{\frac{n_b}{n_a}\right\} && (77)\\ &= \pi/2 && (78) \end{aligned}$$

This relatively-small loss for the sign detector in Gaussian noise is perhaps surprising. The factor $\pi/2 = 1.57 \approx 2dB$ is rather famous.

4.2 Efficacy

The work done in the last subsection to calculate the ARE of the linear (optimal) detector to the sign-detector in Gaussian noise was, while not complicated, rather lengthy. Fortunately there is another way to find the ARE. Suppose we define

$$\begin{aligned} \mu_n(\theta) &= \mathcal{E}\{T_n(x)|\theta\} && (79)\\ \sigma_n^2(\theta) &= \mathcal{V}\{T_n(x)|\theta\} && (80) \end{aligned}$$

in which we consider the detector using test statistic T_n, with n samples on which to operate. We specify the *regularity conditions*

$$\begin{aligned} \frac{d}{d\theta}\,\mu_n|_{\theta=0} &\equiv \dot{\mu}_n(0) > 0 && (81)\\ \lim_{n\to\infty}\left\{\frac{\dot{\mu}_n(0)}{\sqrt{n}\sigma_n(0)}\right\} &\equiv \sqrt{\xi} > 0 && (82)\\ \lim_{n\to\infty}\left\{\frac{\mu_n(\theta)}{\mu_n(0)}\right\} &= 1 && (83)\\ \lim_{n\to\infty}\left\{\frac{\sigma_n(\theta)}{\sigma_n(0)}\right\} &= 1 && (84)\\ T_n(x) & \quad \text{obeys the CLT} && (85) \end{aligned}$$

where

$$\theta = \gamma/\sqrt{n} \tag{86}$$

The quantity ξ is the *efficacy* of the detector.

4.3 The Pitman-Noether Theorem

Suppose in our hypothesis $\theta_l = \gamma/\sqrt{l}$, and that our test $\delta(x)$ uses n samples. We will relate

$$n \ = \kappa(l)^2 l \tag{87}$$

which is general.

Let us examine what happens as $l, n \to \infty$. Under the fifth assumptions of efficacy, we have

$$\alpha \ = \ Q\left(\frac{\tau - \mu_n(0)}{\sigma_n(0)}\right) \tag{88}$$

$$\beta \ = \ Q\left(\frac{\tau - \mu_n(\theta_l)}{\sigma_n(\theta_l)}\right) \tag{89}$$

We solve (88) for the decision threshold τ and insert that to (89) to get

$$\beta \ = \ Q\left(\frac{\sigma_n(\theta)Q^{-1}(\alpha) - (\mu_n(\theta_l) - \mu_n(0))}{\sigma_n(\theta_l)}\right) \tag{90}$$

From the fourth assumption we have

$$\beta \ = \ Q\left(Q^{-1}(\alpha) - \frac{\mu_n(\theta_l) - \mu_n(0)}{\sigma_n(\theta_l)}\right) \tag{91}$$

From the first and third assumption we get

$$\beta \ = \ Q\left(Q^{-1}(\alpha) - \frac{\theta_l \dot{\mu}_n(0)}{\sigma_n(\theta_l)}\right) \tag{92}$$

$$= \ Q\left(Q^{-1}(\alpha) - \frac{\theta_l \sqrt{n}\dot{\mu}_n(0)}{\sqrt{n}\sigma_n(\theta_l)}\right) \tag{93}$$

$$= \ Q\left(Q^{-1}(\alpha) - \frac{\gamma\sqrt{n}}{\sqrt{l}}\frac{\dot{\mu}_n(0)}{\sqrt{n}\sigma_n(\theta_l)}\right) \tag{94}$$

$$= \ Q\left(Q^{-1}(\alpha) - \frac{\gamma\sqrt{n}}{\sqrt{l}}\frac{\dot{\mu}_n(0)}{\sqrt{n}\sigma_n(0)}\right) \tag{95}$$

$$= \ Q\left(Q^{-1}(\alpha) - \frac{\gamma\sqrt{n}}{\sqrt{l}}\sqrt{\xi}\right) \tag{96}$$

$$= \ Q\left(Q^{-1}(\alpha) - \gamma\kappa\sqrt{\xi}\right) \tag{97}$$

where in the fourth line we have used the fourth efficacy assumption (again), in the penultimate line we have used the second efficacy assumption – its definition – and in the last we have used (87).

To compare two detectors (δ_a & δ_b) that have the same ROC operating point we must therefore have

$$\kappa_a \sqrt{\xi}_a \;=\; \kappa_b \sqrt{\xi}_b \tag{98}$$

Consequently

$$\begin{aligned} \mathrm{ARE}_{a,b} &= \lim_{l\to\infty} \left\{ \frac{n_b}{n_a} \right\} && (99) \\ &= \lim_{l\to\infty} \left\{ \frac{\kappa_b^2}{\kappa_a^2} \right\} && (100) \\ &= \frac{\xi_a}{\xi_b} && (101) \end{aligned}$$

This is the Pitman-Noether Theorem: the ARE is the ratio of efficacies.

5 More about Efficacy

5.1 Performance Prediction

The efficacy is a form of *asymptotic* signal to noise ratio. To see this, examine (96) and write it as

$$\begin{aligned} \beta &= Q\left(Q^{-1}(\alpha) - \frac{\gamma\sqrt{n}}{\sqrt{l}} \sqrt{\xi} \right) && (102) \\ &= Q\left(Q^{-1}(\alpha) - \theta\sqrt{n}\sqrt{\xi} \right) && (103) \end{aligned}$$

That is, if the signal (θ) is sufficiently small and the detection record n is sufficiently long, then (103) can plot the approximate ROC.

5.2 Nonlinear Correlator Efficacy

As we have seen before, the use of

$$T(x) \;=\; \sum_{i=1}^{n} a_i g(x_i) \tag{104}$$

makes sense in testing situations like

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \\ \mathcal{K}: &\quad x_i \;=\; \theta s_i \;+\; \nu_i \end{aligned} \tag{105}$$

when θ is small and the noise $\{\nu_i\}$ is iid with pdf $f(x)$. We will assume that

$$\int g(x)f(x)dx \;=\; 0 \tag{106}$$

Optimally we should have $a_i = s_i$ and $g(x) = g_{lo}(x)$, but let us remain general. Define

$$P_s^2 \;\equiv\; \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} s_i^2 \tag{107}$$

$$P_a^2 \;\equiv\; \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} a_i^2 \tag{108}$$

$$P_{as} \;\equiv\; \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} a_i s_i \tag{109}$$

We have

$$\mathcal{E}\{T|\theta\} \;=\; \int\left(\sum_{i=1}^{n} a_i g(x_i)\prod_{i=1}^{n} f(x-\theta s_i)dx\right) \tag{110}$$

$$\;=\; \sum_{i=1}^{n} a_i\int g(x)f(x-\theta s_i)dx \tag{111}$$

and hence

$$\frac{d}{d\theta}\mathcal{E}\{T|\theta\} \;=\; \sum_{i=1}^{n} a_i\int g(x)\left(-s_i\dot{f}(x-\theta s_i)\right)dx \tag{112}$$

$$\frac{d}{d\theta}\mathcal{E}\{T|\theta\}\Big|_{\theta=0} \;=\; \sum_{i=1}^{n}(-s_i a_i)\int g(x)\dot{f}(x)dx \tag{113}$$

Since we trivially have

$$\mathcal{V}\{T|\theta\}|_{\theta=0} \;=\; \sum_{i=1}^{n} s_i^2\int g(x)^2 f(x)dx \tag{114}$$

we can calculate the efficacy as

$$\xi \;=\; \frac{P_{as}}{P_s^2}E(g,f)^2 \tag{115}$$

where[3]

$$E(g,f) \;\equiv\; \frac{\int g(x)\dot{f}(x)dx}{\sqrt{\int g(x)^2 f(x)dx}} \tag{116}$$

[3]I apologize for the notation, this is to be in common with Kassam's text

5.3 Optimal Efficacy

From (115) we see from the Schwarz Inequality that

$$\begin{aligned} P_{as}^2 &= \left(\frac{1}{n}\sum_{i=1}^{n} a_i s_i\right)^2 && (117)\\ &\leq \left(\frac{1}{n}\sum_{i=1}^{n} a_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} s_i^2\right) && (118)\\ &= P_a^2 P_s^2 && (119)\end{aligned}$$

with equality iff $a_i = s_i$. This is perhaps not surprising: you should correlate against the true signal if you have it.

But the Schwarz Inequality also gives us

$$\begin{aligned} \left(\int g(x)\dot{f}(x)dx\right)^2 &= \left(\int g(x)\sqrt{f(x)}\frac{\dot{f}(x)}{\sqrt{f(x)}}dx\right)^2 && (120)\\ &\leq \left(\int g(x)^2 f(x)dx\right)\left(\int \frac{\dot{f}(x)^2}{f(x)}dx\right) && (121)\end{aligned}$$

or

$$E(g,f)^2 \leq \left(\int \frac{\dot{f}(x)^2}{f(x)}dx\right) = I(f) \qquad (122)$$

with equality iff

$$g(x) = \frac{\dot{f}(x)^2}{f(x)} = g_{lo}(x) \qquad (123)$$

Again that the locally-optimal nonlinearity maximizes efficacy is not so surprising. But perhaps the appearance of *Fisher's Information* in (122) is at least interesting.

5.4 Special Cases

In the *linear correlator* case that $g(x) = x$, we have

$$\begin{aligned} \int g(x)\dot{f}(x)dx &= -\int \dot{g}(x)f(x)dx && (124)\\ &= -\int f(x)dx && (125)\\ &= -1 && (126)\end{aligned}$$

and

$$\int g(x)^2 f(x)dx = \int x^2 f(x)dx \quad (127)$$

$$= \sigma^2 \quad (128)$$

This means that for the linear-correlator

$$\xi = P_{as}/\sigma^2 \quad (129)$$

regardless of the noise density.

In the *sign correlator* case that $g(x) = \text{sgn}(x)$, we have

$$\int g(x)\dot{f}(x)dx = -\int \dot{g}(x)f(x)dx \quad (130)$$

$$= -\int 2\delta(x)f(x)dx \quad (131)$$

$$= 2f(0) \quad (132)$$

and

$$\int g(x)^2 f(x)dx = \int \text{sgn}(x)^2 f(x)dx \quad (133)$$

$$= 1 \quad (134)$$

This means that for the sign-correlator

$$\xi = P_{as} 4f(0)^2 \quad (135)$$

regardless of the noise density. In the case that the noise is Gaussian

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (136)$$

we get

$$\xi = P_{as} \frac{2}{\pi\sigma^2} \quad (137)$$

by evaluating at zero. Taking the ratio of the efficacy of the sign-correlator to that of the linear correlator, we easily get 2π – as we found before with much greater difficulty.

ECE 6124
Signal Detection
Noise Models and Detection Structures

Peter Willett

Fall 2018

1 Parametric Models

1.1 Heavy-Tailed Noise

It would be unusual to find a Gaussian random variable more than three standard deviations away from its mean. This benign behavior is why the linear correlator is such a favorable structure: a deviation from mean is meaningful and should add its weight to the decision. Some noise processes, on the other hand, contain outlying samples, these generally from rare impulsive events (lightning, ice-cracks, shrimp-claw snaps) that can easily cause excursions from mean value of ten or more standard deviations. A linear correlator – or any linear processing – of such noise can be catastrophic. In fact, in the case of one of the heaviest-tailed noise distributions that we have considered – Cauchy noise, that does not even *have* a standard deviation[1] – large observations are suppressed from influence on the decision.

At this point it is worthwhile to mention a statistic that at least to some extent measures *heavy-tailed*ness[2] Such statistics can be useful, since a distributional plot (a histogram, if empirical) often appears as bell-shaped, and even a Cauchy distribution can appear little different from Gaussian unless plotted logarithmically. At any rate, the statistic of which we speak is the *kurtosis*, defined for a zero-mean random-variable as

$$\kappa \equiv \frac{\mathcal{E}\{x^4\}}{\mathcal{E}\{x^2\}^2} \tag{1}$$

Since for any combination of jointly-Gaussian random variables $\{x_i\}_{i=1}^4$ we

[1] The Cauchy *pdf* does not even have a *mean*: the integral does not converge.

[2] The "tail" of the distribution is the amount of probability that is many standard deviations away from the mean. A distribution is referred to as heavy-tailed when that weight is significantly larger than Gaussian.

can write

$$\mathcal{E}\{x_1x_2x_3x_4\} \;=\; \mathcal{E}\{x_1x_2\}\mathcal{E}\{x_3x_4\} \;+\; \mathcal{E}\{x_1x_3\}\mathcal{E}\{x_2x_4\} \;+\; \mathcal{E}\{x_1x_4\}\mathcal{E}\{x_2x_3\} \tag{2}$$

we can take all of them to be x and see that for a Gaussian random variable $\kappa = 3$. (Some authors normalize the kurtosis by dividing it by 3.)

Now we can use the generalize Markov inequality to write for any random variable x

$$\begin{aligned} Pr(|x-\mu| > \varepsilon\sigma) &= \mathcal{E}\{\mathcal{I}(|x-\mu| > \varepsilon\sigma)\} && (3)\\ &\leq \mathcal{E}\left\{\left(\frac{x}{\varepsilon\sigma}\right)^4\right\} && (4)\\ &= \kappa/\varepsilon^4 && (5) \end{aligned}$$

This development is perhaps familiar to some as identical to the way the Chebychev inequality is justified; and indeed it can be applied to any absolute moment, and in fact we shall see it again in the context of the Chernoff bound. The implication here is that the kurtosis does have some relevance to measurement of tail weight.

1.2 Gaussian Mixture Noise

Since heavy-tailedness can be caused by rare outlying sample. As such, it is reasonable to consider a two-element Gaussian mixture

$$f(x) \;=\; (1-\varepsilon)\frac{1}{\sqrt{2\pi}\sigma_1}e^{-x^2/2\sigma_1^2} \;+\; \varepsilon\frac{1}{\sqrt{2\pi}\sigma_2}e^{-x^2/2\sigma_2^2} \tag{6}$$

with the idea that the former term is "ambient" well-behaved noise and the latter the impulsive – ε is typically very small. It is common practice to use moment-matching to estimate the parameters, and since there are three parameters three moments are needed:

$$\begin{aligned} \mathcal{E}[\{x^2\}] &= (1-\varepsilon)\sigma_1^2 + \varepsilon\sigma_2^2 && (7)\\ \mathcal{E}[\{x^4\}] &= (1-\varepsilon)3\sigma_1^4 + \varepsilon 3\sigma_2^4 && (8)\\ \mathcal{E}[\{x^6\}] &= (1-\varepsilon)15\sigma_1^6 + \varepsilon 15\sigma_2^6 && (9) \end{aligned}$$

Please see figures 1 and 2 for a plot of the Gaussian-mixture *pdf*. The locally-optimal nonlinearity for a Gaussian-mixture is

$$g_{lo}(x) \;=\; \frac{(1-\varepsilon)\frac{x}{\sigma_1^3}e^{-x^2/2\sigma_1^2} \;+\; \varepsilon\frac{x}{\sigma_2^3}e^{-x^2/2\sigma_2^2}}{(1-\varepsilon)\frac{1}{\sigma_1}e^{-x^2/2\sigma_1^2} \;+\; \varepsilon\frac{1}{\sigma_2}e^{-x^2/2\sigma_2^2}} \tag{10}$$

and this is plotted in figure 3.

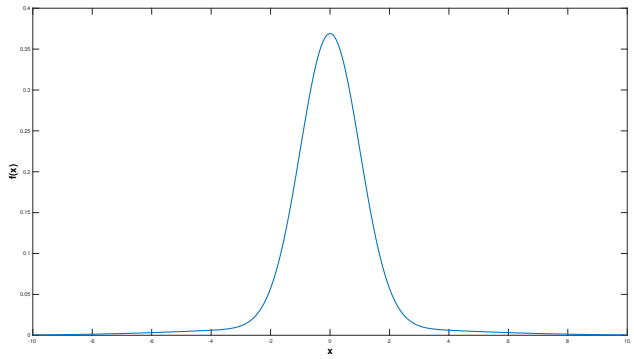

Figure 1: Gaussian mixture with $\varepsilon = 0.1$ (larger than it would normally be), $\sigma_1 = 1$ and $\sigma_2 = 4$ (smaller than normal). The values chosen are for graphical purposes, to emphasize the behavior. Note that the *pdf* does not, by eye, seem to differ much from Gaussian.

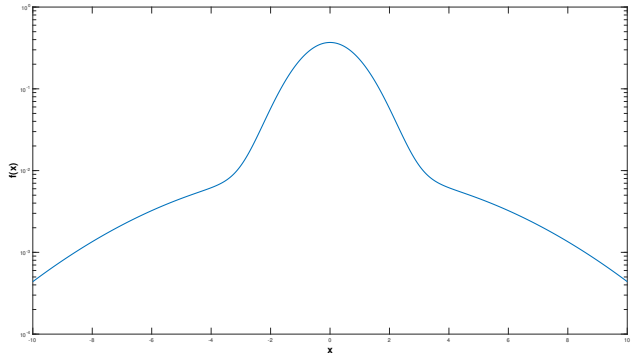

Figure 2: Log plot corresponding to figure 1. The non-Gaussian behavior is obvious now, and takes the form of a "knee".

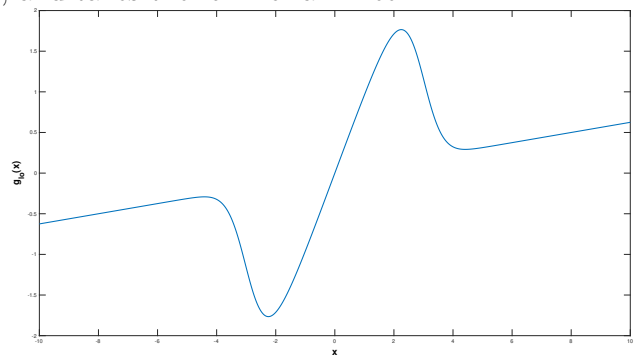

Figure 3: Locally-optimal nonlinearity corresponding to figure 1.

1.3 Generalized Gaussian Noise

Generalized Gaussian noise has *pdf*

$$f(x) \;=\; ce^{-|x/a|^k} \tag{11}$$

Given that $k > 0$ we can evaluate

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx && (12)\\ &= c\int_{-\infty}^{\infty} e^{-|x/a|^k}dx && (13)\\ &= 2c\int_{0}^{\infty} e^{-|x/a|^k}dx && (14)\\ &= 2ca/k\int_{0}^{\infty} t^{1/k-1}e^{-t}dt && (15)\\ &= (2ca/k)\,\Gamma(1/k) && (16) \end{aligned}$$

where the substitution $t = (x/a)^k$ is used; or

$$c \;=\; \frac{k}{2a\Gamma(1/k)} \tag{17}$$

In a similar way[3] we could find

$$\begin{aligned} \sigma^2 &= \frac{a^2\Gamma(3/k)}{\Gamma(1/k)} && (18)\\ \kappa &= \frac{\Gamma(5/k)\Gamma(1/k)}{\Gamma(3/k)^2} && (19)\\ I(f) &= \frac{k^2\Gamma(3/k)\Gamma(2-1/k)}{\sigma^2\Gamma(1/k)^2} && (20) \end{aligned}$$

And since

$$g_{lo}(x) \;=\; \frac{d}{dx}(\log(f(x)) \tag{21}$$

we can write

$$g_{lo}(x) \;\propto\; \mathrm{sgn}(x)|x|^{k-1} \tag{22}$$

Note that the cases $k = 1$ and $k = 2$ correspond, respectively, to Laplace and Gaussian noises – the sign-detector and linear correlator follow.

[3]For reference, for b even we have $\mathcal{E}\{x^b\} = a^b\Gamma\left(\frac{b+1}{k}\right) / \left(\frac{1}{k}\right)$.

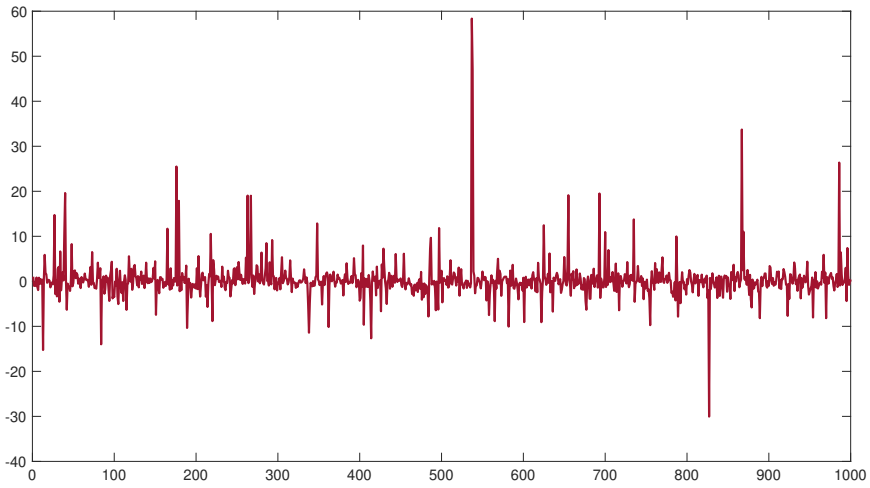

Figure 4: Johnson noise with $\delta = 0.75$.

1.4 Johnson Noise

There is a broad class of *transformation noise*, for which

$$x = g^{-1}(u) \tag{23}$$

the underlying random variable u is unit-normal, and $g(\cdot)$ is an invertible nonlinearity. It is straightforward to see

$$\begin{aligned} F(x) &= Pr(g^{-1}(u) < x) && (24)\\ &= Pr(u < g(x)) && (25)\\ &= 1 - Q(g(x)) && (26)\\ f(x) &= \dot{g}(x)\frac{1}{\sqrt{2\pi}}e^{-g(x)^2/2} && (27) \end{aligned}$$

For the particular transformation known as Johnson noise, we have

$$g^{-1}(u) = \lambda \sinh(u/\delta) \tag{28}$$

The parameters δ and λ can be inferred from moments[4] as

$$\begin{aligned} \lambda &= \sqrt{\frac{2\sigma^2}{\sqrt{2\kappa - 2} - 2}} && (29)\\ \delta &= \sqrt{\frac{2}{\log(\sqrt{2\kappa - 2} - 1)}} && (30) \end{aligned}$$

[4]The method of moments is easy; however, maximum likelihood estimation of the parameters may give a slightly better fit.

Note that one can write

$$x \;=\; \frac{\lambda}{2}(e^{u/\delta} - e^{-u/\delta}) \tag{31}$$

so

$$e^{2u/\delta} \;-\; \frac{2}{\lambda}e^{u/\delta} - 1 \;=\; 0 \tag{32}$$

hence by the quadratic formula

$$g(x) \;=\; \delta \log\left(\frac{x}{\lambda} + \sqrt{\left(\frac{x}{\lambda}\right)^2 + 1}\right) \tag{33}$$

$$\dot{g}(x) \;=\; \frac{\delta}{\sqrt{\lambda\left(\frac{x}{\lambda}\right)^2 + 1}} \tag{34}$$

The formulas for $f(x)$ and $g_{lo}(x)$ follow but are tedious.

It turns out that Johnson noise has some characteristics of log-normality (substitute (33) to the exponent of the Gaussian). What is perhaps appealing about Johnson noise is that it needs only two parameters to match a heavy-tailed *pdf*. An example noise trace is in figure 4.

1.5 Alpha-Stable Noise

Consider that we are interested in a random variable with characteristic function

$$\phi(\omega) \;=\; e^{-|\beta\omega|^\alpha} \tag{35}$$

This concept attracted some attention a few years ago, because in keeping with the exponential nature of $\phi(\omega)$ such random variables have the property that when they are added together the sum remains in the same family (only β will change). It is also nice in that $\alpha = 2$ and $\alpha = 1$ are respectively Gaussian and Cauchy. However, since other choices for α do not have explicit *pdf*'s (they are represented as infinite series), this nice fact loses some of its attraction.

Additionally, considering that except for Gaussian random variables one should avoid using linear operations, the α-stable family loses essentially all other appeal. It would be interesting to find a non-Gaussian density $f(x)$ such that if $g_{lo}(x)$ has *pdf* $\tilde{f}(\cdot)$, then sums of $g_{lo}(x_i)$'s remained in the family $\tilde{f}(\cdot)$. As far as I am aware no one has attacked this question, and there may be no such $f(\cdot)$.

1.6 Series Expansions

Modulo situations of pathology, any *pdf* can be represented via a series. An especially favorable sort of series is

$$f(x) \;=\; \hat{f}(x) \sum_{i=1}^{\infty} c_i \psi_i(x) \tag{36}$$

in which

$$\int \hat{f}(x) \psi_i(x) \psi_j(x) dx \;=\; \delta_{i=j} \tag{37}$$

and $\hat{f}$ is some canonical *pdf* (like Gaussian). The choice of $\hat{f}$ actually matters little; but convergence is better if f and $\hat{f}$ are "close". It is easy to show that

$$c_i \;=\; \int \psi_i(x) f(x) dx \;=\; \mathcal{E}\{\psi_i(x)\} \tag{38}$$

to determine the coefficients.

One example is $\hat{f}(x) = \phi(x)$ (the Gaussian *pdf*), for which

$$\psi_i(x) \;=\; \frac{1}{\sqrt{i!}} \frac{\phi^{(i)}(x)}{\phi(x)} \tag{39}$$

and the superscript (i) denotes the i^{th} derivative. If $\hat{f}$ is exponential the functions are Laguerre polynomials; if uniform they are Jacobi.

2 Bivariate Densities

2.1 Series Expansions

Again, modulo situations of pathology, any bivariate *pdf* can be represented via a series

$$f(x,y) \;=\; \hat{f}(x)\hat{f}(y) \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} c_{i,j} \psi_i(x) \psi_j(y) \tag{40}$$

Special cases include the diagonal

$$f(x,y) \;=\; \hat{f}(x)\hat{f}(y) \sum_{i=1}^{\infty} c_i \psi_i(x) \psi_i(y) \tag{41}$$

and the Frechet

$$f(x,y) \;=\; \hat{f}(x)\hat{f}(y)(1 + c\psi(x)\psi(y)) \tag{42}$$

which can be shown only to be possible for (hidden-) Markov switched system.

2.2 Poor's Locally-Optimal Memoryless Detector

Poor & Thomas asked what would happen in the standard detection situation that a known and vanishingly-small signal in noise

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \theta \;+\; \nu_i \end{aligned} \tag{43}$$

was to be detected via nonlinear correlator

$$T(x) \;=\; \sum_{i=1}^{n} g(x_i) \tag{44}$$

The wrinkle is that the correlator is *memoryless* but the noise is not independent[5]. One might at first glance expect the best $g(x)$ to be $g_{lo}(x)$, but that turns out not to be the case. Assuming $\mathcal{E}\{g(x)|\mathcal{H}\} = 0$, we have

$$\frac{d}{d\theta}\,\mathcal{E}\{T(x)\}|_{\theta=0} \;=\; n\int \dot{f}(x)g(x)dx \tag{45}$$

We also have

$$\mathcal{V}\{T(x)\}|_{\theta=0} \quad = \quad \sum_{i=1}^{n}\sum_{j=1}^{n}\int g(x_i)g(x_j)f_{i-j}(x_i,x_j)dx_i dx_j \tag{46}$$

$$\rightarrow \quad n\int f(x)g(x)^2dx \;+\; n\int g(x)g(y)K(x,y)dxdy \tag{47}$$

where the $\rightarrow$ indicates that this becomes equality as n diverges and

$$K(x,y) \;\equiv\; \sum_{k=1}^{n}(f_k(x,y) + f_{-k}(x,y) \tag{48}$$

Now we have

$$\xi \;=\; \frac{(n\int \dot{f}(x)g(x)dx)^2}{n(n\int f(x)g(x)^2dx \;+\; n\int g(x)g(y)K(x,y)dxdy)} \tag{49}$$

as the efficacy.

We want to maximize this, so we set up the proxy functional

$$J(g(x)) \;\equiv\; \int \dot{f}(x)g(x)dx + \lambda\left(\int f(x)g(x)^2dx \;+\; \int g(x)g(y)K(x,y)dxdy\right) \tag{50}$$

[5]We will assume wide-sense stationarity and a certain condition called "strong mixing" that basically says that the central limit theorem will apply to $T(x)$.

We apply *calculus of variations*, and insist that at an optimum $g(x)$ we must have

$$\frac{d}{d\epsilon} J(g(x) + \epsilon\delta(x))|_{\epsilon=0} \;=\; 0 \tag{51}$$

irrespective of $\delta(x)$. The argument is that if

$$\frac{d}{d\epsilon} J(g(x) + \epsilon\delta(x))|_{\epsilon=0} \;>\; 0 \tag{52}$$

then clearly $g(x)+\epsilon\delta(x)$ is better than $g(x)$ for some small ϵ. However, even if we have

$$\frac{d}{d\epsilon} J(g(x) + \epsilon\delta(x))|_{\epsilon=0} \;<\; 0 \tag{53}$$

this implies that $g(x)-\epsilon\delta(x)$ is better than $g(x)$ for some small ϵ, and hence even then $g(x)$ could not be optimum.

Hence, we write

$$\begin{aligned} J(g(x)) \;&=\; \int \dot{f}(x)[g(x) + \epsilon\delta(x)]dx \;+\; \lambda\left(\int f(x)[g(x) + \epsilon\delta(x)]^2 dx\right. \\ &\left. + \int [g(x) + \epsilon\delta(x)][g(y) + \epsilon\delta(y)]K(x,y)dxdy\right) \end{aligned} \tag{54}$$

so

$$\begin{aligned} \frac{d}{d\epsilon} J(g(x) + \epsilon\delta(x))|_{\epsilon=0} \;&=\; \int \dot{f}(x)\delta(x)dx \;+\; \lambda\left(2\int f(x)\delta(x)g(x)dx\right. \\ &\left. + \int [\delta(x)g(x) + \delta(y)g(y)]K(x,y)dxdy\right) \end{aligned} \tag{55}$$

To make this zero irrespective of $\delta(x)$ we require

$$\dot{f}(x) + 2\lambda g(x)f(x) + 2\lambda \int K(x,y)g(y)dy \;=\; 0 \tag{56}$$

or

$$\frac{\dot{f}(x)}{f(x)} + 2\lambda g(x) + 2\lambda \int \frac{K(x,y)}{f(x)} g(y)dy \;=\; 0 \tag{57}$$

Realizing that λ is a constant to be determined, but that the constant contributes only as a scalar multiplier on $g(x)$, we write

$$g_{lo}(x_i) \;=\; g(x_i) \;-\; \mathcal{E}\left\{g_{lo}(x_i)|\{x_j\}_{i\neq j}\right\} \tag{58}$$

which makes sense and is rather beautiful. It can easily be shown that the expectation is zero if the noise is independent; and that it is linear (i.e., a full matched filter) if the noise is Gaussian.

ECE 6124
Signal Detection
Locally Optimal Quantized Detection

Peter Willett

Fall 2018

1 Introduction

A quantized detection system maps each observation to one of L levels prior to taking its place in a decision. Such a scheme may be representative of *decentralized detection*, in which the quantization is performed to save bandwidth before transmission from local sensor to data *fusion* center. For example, if $L = 8$ only 3 bits are needed for each transmission; and we have seen that even for a single bit ($L = 2$) only about $2dB$ is lost relative to optimal in an additive Gaussian situation. Please also note that "3 bits" is actually very conservative, since a data compression system could bring this down to a very low level indeed. Another important note is that while in the following we will be at pains to determine the optimal quantization scheme – both thresholds $\{t_k\}_{k=0}^{L}$ and levels $\{l_k\}_{k=0}^{L-1}$ – we do not need to transmit between local sensor and fusion center at this level, but instead simply need to tell the fusion center that if it gets (decodes?) message k then it should use level l_k.

There are two schemes on how to perform such quantization. In the first, we quantize the data x_i directly. Some thought tells us that such quantization amounts to use of a *staircase*-style nonlinearity $g(x)$ in a normal correlator/detector structure. We will gain insight from this; but that insight will suggest – and we shall show – that, in fact, it is better to use a nonlinearity (actually $g_{lo}(x)$) first, and quantize later.

We shall perform our analysis in the small-signal regime: locally-optimal quantization and performance measured by efficacy, and an additive known signal. In fact the results can be shown to extend much farther, to cases of non-vanishing signals: it can be shown that a likelihood-ratio quantization is optimal, given independence over i when conditioned on the hypothesis. This also implies that data can be accumulated prior to quantization, meaning that the likelihood ratio can formed of integrated data at each sensor.

2 Direct Data (Abscissa) Quantization

2.1 The Nonlinearity

Consider that we have thresholds $-\infty = t_0 < t_1 < \ldots < t_{L-1} < t_L = \infty$ and we define quantization regions

$$A_k \;=\; \{x : t_k \leq x < t_{k+1}\}, \quad k = 0, 1, \ldots (L-1) \tag{1}$$

which we will assume here are identical[1] for each observation/sensor[2].

$$\begin{aligned} p_{ki}(\theta) &\equiv Pr(x_i \in A_k|\theta) && (2)\\ &= Pr(t_k \leq x_i < t_{k+1}|\theta) && (3)\\ &= F(t_k - \theta s_i) \;-\; F(t_{k+1} - \theta s_i) && (4) \end{aligned}$$

Let us also define

$$z_{ik} \;\equiv\; \left\{ \begin{array}{ll} 1 & x_i \in A_k \\ 0 & x_i \notin A_k \end{array} \right\} \tag{5}$$

If it clear that $\mathbf{z} = \{z_{ik}\}$ determines the quantized observations, albeit rather profligately. But we can use $\mathbf{z}$ to write the likelihood of the observation *after quantization* – the $\mathbf{z}$ is what the fusion center / decision-maker will use to make its final decision. We have

$$f(\mathbf{z}) \;=\; \prod_{i=1}^{n} \prod_{k=0}^{L-1} p_{ki}(\theta)^{z_{ik}} \tag{6}$$

Thence we write

$$\begin{aligned} \log(f(\mathbf{z})) &= \sum_{i=1}^{n} \sum_{k=0}^{L-1} z_{ik} \log(p_{ki}(\theta)) && (7)\\ \frac{d}{d\theta} \log(f(\mathbf{z}))|_{\theta=0} &= \sum_{i=1}^{n} \sum_{k=0}^{L-1} z_{ik} \frac{\dot{p}_{ki}(0)}{p_{ki}(0)} && (8)\\ &= \sum_{i=1}^{n} \sum_{k=0}^{L-1} z_{ik} s_i \frac{f(t_{k+1}) - f(t_k)}{F(t_{k+1}) - F(t_k)} && (9) \end{aligned}$$

[1] We shall see shortly that the quantizers ought to be identical if the noises at the sensors are the same.

[2] We will use sensor and observation interchangeably. Most of our previous work has assumed time-series detection, hence "observation" seems appropriate; but quantization probably makes the most sense in the context of data fusion, where we should speak of "sensors." To keep commonality with previous work, we will probably have time series detection as our mental model.

$$= \sum_{i=1}^{n} s_i q(x_i) \tag{10}$$

$$\tag{11}$$

in which

$$q(x) \equiv l_k \text{ when } t_k \le x < t_{k+1} \tag{12}$$

and

$$l_k = \frac{f(t_{k+1}) - f(t_k)}{F(t_{k+1}) - F(t_k)} \tag{13}$$

An example is plotted in figure 1, and the message is that this "quantizer" is really just a nonlinearity, similar to what we always have with nonlinear correlators.

It is interesting to evaluate

$$\mathcal{E}\{g_{lo}(x) | t_k \le x < t_{k+1}\} = \frac{\int_{t_k}^{t_{k+1}} \frac{-\dot{f}(x)}{f} x f(x) dx}{\int_k^{t_{k+1}} f(x) dx} \tag{14}$$

$$= \frac{f(t_{k+1}) - f(t_k)}{F(t_{k+1}) - F(t_k)} \tag{15}$$

$$= l_k \tag{16}$$

which has intuitive appeal.

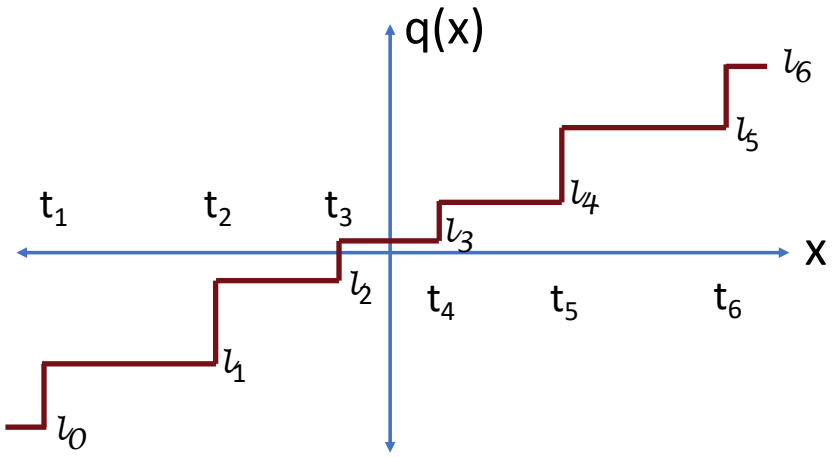

Figure 1: Notional figure of a quantizer – that is, a nonlinearity that happens to have a "staircase" form.

2.2 Optimizing the Nonlinearity

We are interested in

$$E(q,f)^2 = \frac{(\int q\dot{f})^2}{\int q^2 f} \quad (17)$$

$$= \frac{(\int \dot{q}f)^2}{\int q^2 f} \quad (18)$$

Now since

$$\dot{q}(x) = \sum_{k=1}^{L-1}(l_k - l_{k-1})\delta(x - t_k) \quad (19)$$

we have

$$\int \dot{q}f = \sum_{k=1}^{L-1}(l_k - l_{k-1})f(t_k) \quad (20)$$

$$= \sum_{k=1}^{L-1} l_k(f(t_{k+1}) - f(t_k)) \quad (21)$$

$$= \sum_{k=1}^{L-1} \frac{(f(t_{k+1}) - f(t_k))^2}{F(t_{k+1}) - F(t_k)} \quad (22)$$

We also have

$$\int q^2 f = \sum_{k=1}^{L-1} l_k^2(f(t_{k+1}) - f(t_k)) \quad (23)$$

$$= \sum_{k=1}^{L-1} \frac{(f(t_{k+1}) - f(t_k))^2}{F(t_{k+1}) - F(t_k)} \quad (24)$$

in which we have assumed, without loss of generality, that we have

$$\int q(x)f(x)dx = 0 \quad (25)$$

Putting these together, we have

$$E(q,f)^2 = \sum_{k=1}^{L-1} \frac{(f(t_{k+1}) - f(t_k))^2}{F(t_{k+1}) - F(t_k)} \quad (26)$$

as something we need to optimize over $\{t_k\}$.

Let us differentiate $E(q,f)^2$. We have

$$\begin{aligned}
0 &= \frac{d}{dt_k}E(q,f)^2 && (27)\\
&= 2\dot{f}(t_k)\frac{f(t_k)-f(t_{k-1})}{F(t_k)-F(t_{k-1}} - f(t_k)\frac{(f(t_k)-f(t_{k-1}))^2}{(F(t_k)-F(t_{k-1}))^2} && (28)\\
&\quad -2\dot{f}(t_k)\frac{f(t_{k+1})-f(t_k)}{F(t_{k+1})-F(t_k)} + f(t_k)\frac{(f(t_{k+1})-f(t_k))^2}{(F(t_{k+1})-F(t_k))^2} && (29)\\
&= 2\dot{f}(t_k)l_k - f(t_k)l_k^2 - \dot{f}(t_k)l_{k+1} + f(t_k)l_{k+1}^2 && (30)\\
0 &= 2g_{lo}(t_k)(l_k - l_{k+1}) - (l_k^2 - l_{k+1}^2) && (31)\\
0 &= 2g_{lo}(t_k) - (l_k + l_{k+1}) && (32)
\end{aligned}$$

which implies that

$$g_{lo}(t_k) \;=\; \frac{l_k + l_{k+1}}{2} \qquad (33)$$

Interestingly, back and forth recursion on (16) & (33) generally converges, and is known as the (generalized) *Lloyd-Max* algorithm for quantizer design.

2.3 Minimizing the MSE

Suppose we began with a different problem: we want to select $\{l_k\}$ and $\{t_k\}$ to minimize

$$MSE \;=\; \mathcal{E}\{(q(x) - g_{lo}(x))^2\} \qquad (34)$$

We write

$$MSE \;=\; \sum_{k=0}^{L-1}\int_{t_k}^{t_{k+1}} (l_k - g_{lo}(x)))^2 f(x)dx \qquad (35)$$

Taking the derivative with respect to the levels we get

$$\begin{aligned}
0 &= \frac{d}{t_{l_k}}(MSE) && (36)\\
&= \int_{t_k}^{t_{k+1}} (l_k - g_{lo}(x)))f(x)dx && (37)
\end{aligned}$$

or

$$l_k \;=\; \mathcal{E}\{g_{lo}(x)|t_{k-1} \le x < t_k) \qquad (38)$$

Taking the derivative with respect to the thresholds we find

$$\begin{aligned}
0 &= \frac{d}{t_{t_k}}(MSE) && (39)\\
&= (l_k - g_{lo}(t_k))^2 f(t_k) \;-\; (l_{k+1} - g_{lo}(t_k))^2 f(t_k) && (40)\\
&= (l_k - g_{lo}(t_k))^2 \;-\; (l_{k+1} - g_{lo}(t_k))^2 && (41)\\
&= (l_k + l_{k+1} - 2g_{lo}(t_k))(l_k + l_{k+1}) && (42)
\end{aligned}$$

hence

$$g_{lo}(t_k) \;=\; \frac{l_k + l_{k+1}}{2} \tag{43}$$

Since (38) & (43) are the same as (16) & (33), we conclude that the goals are the same: maximizing the efficacy is the same as minimizing the quantization error between $q(x)$ and $g_{lo}(x)$. If the quantization operation is thought of as introducing noise to $g_{lo}(x)$, this makes sense.

3 Detection-Optimized (Ordinate) Quantization

The previous section developed equations for design of an optimal quantizer whose quantization sets were simply-connected (i.e., convex) sets of input data. That is, the the quantizer operated on the x-axis, the abscissa. In the case of a monotone-increasing locally-optimal nonlinearity $g_{lo}(x)$ that turns out to be the best thing to do. However, (35) suggests that when $g_{lo}(x)$ is not monotone, then there may be better things to do. And there are

First we define the quantization regions as $\{A_k\}$ (disjoint and exhaustive) such that when $x_i \in A_k$ the quantizer reports l_k. Again with (5)

$$z_{ik} \;=\; \left\{ \begin{array}{ll} 1 & x_i \in A_k \\ 0 & x_i \notin A_k \end{array} \right\} \tag{44}$$

the fusion rule (8) still holds:

$$\frac{d}{d\theta} \log(f(\mathbf{z}))|_{\theta=0} \;=\; \sum_{i=1}^{n} \sum_{k=0}^{L-1} z_{ik} \frac{\dot{p}_{ki}(0)}{p_{ki}(0)} \tag{45}$$

where from (2) we have

$$\begin{aligned} p_{ki}(\theta) \;&=\; Pr(x_i \in A_k|\theta) && (46) \\ &=\; \int_{A_k} f(x - \theta s_i) dx && (47) \end{aligned}$$

so

$$\frac{d}{d\theta}\, p_{ki}(\theta)|_{\theta=0} \;=\; -s_i \int_{A_k} \dot{f}(x) dx \tag{48}$$

or

$$\frac{\dot{p}_{ki}(0)}{p_{ki}(0)} \;=\; s_i \mathcal{E}\{g_{lo}(x)|x \in A_k\} \tag{49}$$

and we define $q(x) = l_k$ where $x \in A_k$.

Now, repeating (17), we have

$$E(q,f)^2 \;=\; \frac{(\int q\dot{f})^2}{\int q^2 f} \tag{50}$$

We have

$$\begin{aligned}
\int q\dot{f} &= \sum_{k=0}^{L-1} \int_{A_k} l_k \dot{f}(x)dx && (51)\\
&= \sum_{k=0}^{L-1} -l_k \mathcal{E}\{g_{lo}(x)|x \in A_k\} \left(\int_{A_k} f(x)dx\right) && (52)\\
&= \sum_{k=0}^{L-1} l_k \mathcal{E}\{g_{lo}(x)|x \in A_k\} \left(\int_{A_k} f(x)dx\right) && (53)\\
&= \sum_{k=0}^{L-1} l_k^2 \left(\int_{A_k} f(x)dx\right) && (54)
\end{aligned}$$

We also have, assuming $\int qf = 0$,

$$\begin{aligned}
\int q^2 f &= \sum_{k=0}^{L-1} l_k^2 \int_{A_k} f(x)dx && (55)\\
&= \int q\dot{f} && (56)
\end{aligned}$$

and hence (this generalizes (26)) we have

$$E(q,f)^2 \;=\; \sum_{k=0}^{L-1} l_k^2 \int_{A_k} f(x)dx \tag{57}$$

Our goal now is to optimize the $\{A_k\}$.

First, let us assume we have two sets, A_j & A_k for which $l_j < l_k$. Further, we assume that we have subsets $B_j \subset A_j$ & $B_k \subset A_k$, such that

$$\int_{B_j} f(x)dx \;=\; \int_{B_k} f(x)dx \tag{58}$$

and

$$\mathcal{E}\{g_{lo}(x)|x \in B_j\} \;> \mathcal{E}\{g_{lo}(x)|x \in B_k\} \tag{59}$$

These are illustrated in figure 2. As suggested by the figure, we exchange the two out-of-place subsets

$$\begin{aligned}
\hat{A}_j &= (A_j \bigcap \bar{B}_j) \bigcup B_k && (60)\\
\hat{A}_k &= (A_k \bigcap \bar{B}_k) \bigcup B_j && (61)
\end{aligned}$$

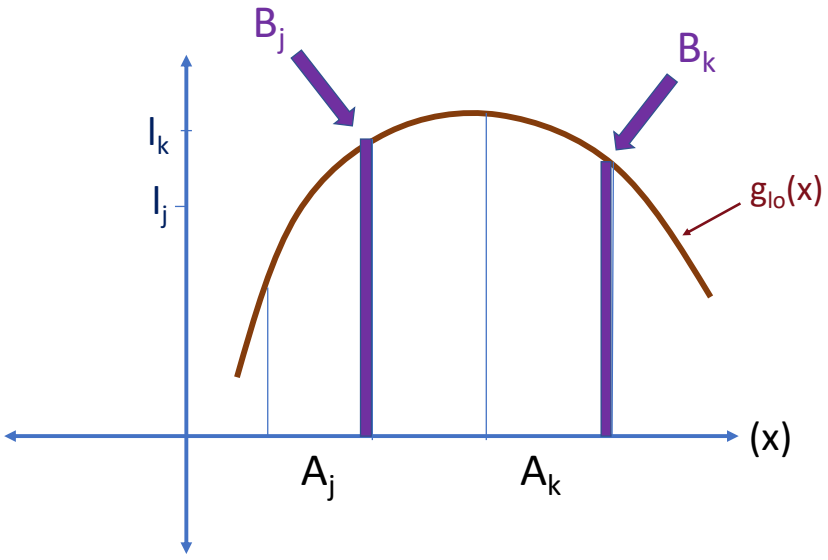

Figure 2: Cartoon suggesting the sets A_j & A_k and their (equal-probability) subsets B_j & B_k. The key is that $l_j < l_k$, but over their subsets $\mathcal{E}\{g_{lo}(x)|x \in B_j\} > \mathcal{E}\{g_{lo}(x)|x \in B_k\}$. The development will switch B_j to A_k and likewise B_k to A_j. Note that no change in $\int_{A_j} f(x)dx$ nor $\int_{A_k} f(x)dx$ results.

where $\bar{C} = \{\forall x \notin C\}$. We have

$$
\begin{aligned}
& E(q,f)^2\Big|_{\{\hat{A}_k\}} - E(q,f)^2\Big|_{\{A_k\}} \\
&= \sum_{k=0}^{L-1} \mathcal{E}\{g_{lo}(x)|x \in \hat{A}_k\}^2 \int_{\hat{A}_k} f(x)dx \qquad (62) \\
&\quad - \sum_{k=0}^{L-1} \mathcal{E}\{g_{lo}(x)|x \in A_k\}^2 \int_{A_k} f(x)dx \\
&= \mathcal{E}\{g_{lo}(x)|x \in \hat{A}_j\}^2 \int_{\hat{A}_j} f(x)dx + \mathcal{E}\{g_{lo}(x)|x \in \hat{A}_k\}^2 \int_{\hat{A}_k} f(x)dx \qquad (63) \\
&\quad - \mathcal{E}\{g_{lo}(x)|x \in A_j\}^2 \int_{A_j} f(x)dx + \mathcal{E}\{g_{lo}(x)|x \in A_k\}^2 \int_{A_k} f(x)dx \\
&= \frac{(\int_{A_j} g_{lo}f - \int_{B_j} g_{lo}f + \int_{B_k} g_{lo}f)^2}{\int_{A_j} f - \int_{B_j} f + \int_{B_k} f} + \frac{(\int_{A_k} g_{lo}f - \int_{B_k} g_{lo}f + \int_{B_j} g_{lo}f)^2}{\int_{A_k} f - \int_{B_k} f + \int_{B_j} f} \\
&\quad - \frac{(\int_{A_j} g_{lo}f)^2}{\int_{A_j} f} - \frac{(\int_{A_k} g_{lo}f)^2}{\int_{A_k} f} \qquad (64)
\end{aligned}
$$

$$= \frac{(\int_{A_j} g_{lo} f - \int_{B_j} g_{lo} f + \int_{B_k} g_{lo} f)^2}{\int_{A_j} f} + \frac{(\int_{A_k} g_{lo} f - \int_{B_k} g_{lo} f + \int_{B_j} g_{lo} f)^2}{\int_{A_k} f}$$

$$- \frac{(\int_{A_j} g_{lo} f)^2}{\int_{A_j} f} - \frac{(\int_{A_k} g_{lo} f)^2}{\int_{A_k} f} \tag{65}$$

$$= 2\left(\int_{B_j} g_{lo} f - \int_{B_k} g_{lo} f\right)\left(\frac{\int_{A_k} g_{lo} f}{\int_{A_k} f} - \frac{\int_{A_j} g_{lo} f}{\int_{A_j} f}\right) \tag{66}$$

$$\left(\int_{B_j} g_{lo} f - \int_{B_k} g_{lo} f\right)^2 \left(\frac{1}{\int_{A_j} f} - \frac{1}{\int_{A_k} f}\right) \tag{67}$$

$$> 2\left(\int_{B_j} g_{lo} f - \int_{B_k} g_{lo} f\right)\left(\frac{\int_{A_k} g_{lo} f}{\int_{A_k} f} - \frac{\int_{A_j} g_{lo} f}{\int_{A_j} f}\right) \tag{68}$$

$$> 0 \tag{69}$$

where the last inequality follows since

$$\int_{B_j} f(x)dx = \int_{B_k} f(x)dx \tag{70}$$

by construction.

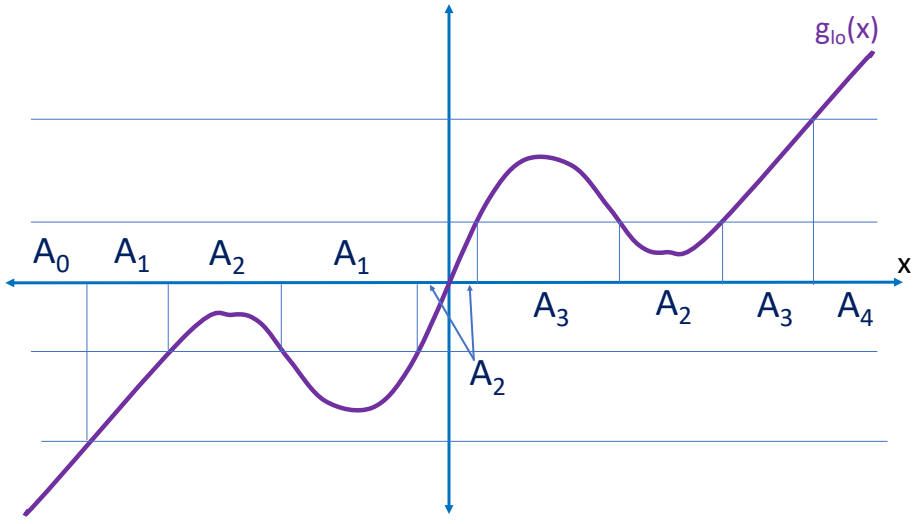

Figure 3: A non-monotone $g_{lo}(x)$ (which looks rather like that corresponding to a Gaussian mixture) and a notional $L = 5$ level quantizer. Note that the quantization regions are not simply-connected in x, but they are in $g_{lo}(x)$.

We therefore have the result that any quantization that is not simply connected in $g_{lo}(x)$ cannot be optimal, and conclude that the proper way to quantize is over $g_{lo}(x)$ – on the y-axis. Naturally, if $g_{lo}(x)$ is monotone, then this is not new. See figure 3 for a notional example of that. In order to design an optimal quantizer it is necessary to transform x to $y = g_{lo}(x)$, and to use the Lloyd-Max procedure in that domain.

Note that this section is not entitled "general" quantization. It is known that an optimal quantizer for conditionally-independent observations / sensors results in simply-connected regions in terms of the local likelihood ratio – this applies to situations in which the hypotheses are not of the additive signal type. Even more interesting, when the observations / sensors are not conditionally-independent. For example, the behavior observed even in the (trivial?) case of a mean shift in correlated Gaussian noise is complex and unexpected.

ECE 6124
Signal Detection
Performance Prediction and Bounds

Peter Willett

Fall 2018

1 Ali-Silvey Distance Measures

Certainly signal-to-noise ratio (SNR) presents an appealing way to measure detection performance. SNR determines the performance exactly in the Gaussian shift-in-mean situation; but for more general additive-signal problems the efficacy, which can be thought of as an asymptotic SNR, is appealing as well. However, it is desirable to propose performance measures applicable in more general and non-asymptotic cases.

One such is the J-divergence

$$J \equiv \mathcal{E}\left\{\log\left(\frac{f(x|\mathcal{K})}{f(x|\mathcal{H})}\right)|\mathcal{K}\right\} - \mathcal{E}\left\{\log\left(\frac{f(x|\mathcal{K})}{f(x|\mathcal{H})}\right)|\mathcal{H}\right\} \tag{1}$$

which has appeal as the distance between the means of the optimal test statistic, the log-likelihood ratio. The J-divergence is the difference between the two *Kullback-Leibler* numbers, and sometime just the second term

$$d_{KL} \equiv \mathcal{E}\left\{\log\left(\frac{f(x|\mathcal{H})}{f(x|\mathcal{K}}\right)|\mathcal{H}\right\} \tag{2}$$

is reported and called the *divergence* or KL divergence. KL divergence has much attention in information theory.

Actually the J-divergence is one example of an *Ali-Silvey* distance measure, for which we have

$$d \equiv h(\mathcal{E}\{C(L)\}) \tag{3}$$

in which $h(\cdot)$ is a monotone-increasing function, L is the likelihood ratio and $C(\cdot)$ is a convex (convec-cup) function. For J-divergence we have

$$h(y) = y \qquad C(y) = (y-1)\log(y) \tag{4}$$

Many other possibilities exist; for example if we have

$$h(y) = -\log(1-y/2) \qquad C(y) = (\sqrt{y}-1)^2 \tag{5}$$

we have the Bhattacharyya distance, which we shall see more about soon.

2 The Chernoff Bound

The Chernoff idea is as follows. Consider a test based on $T(x)$ and using threshold τ.

$$\begin{aligned} P_f &= \Pr(\text{false alarm}) && (6)\\ &= \int \mathcal{I}(T(x) > \tau) f(x|\mathcal{H}) dx && (7)\\ &= \mathcal{E}\{\mathcal{I}(T(x) > \tau)|\mathcal{H}\} && (8)\\ &\leq \mathcal{E}\{e^{s(T(x)-\tau)}|\mathcal{H}\} && (9)\\ &= e^{-s\tau}\mathcal{E}\{e^{T(x)}|\mathcal{H}\} && (10) \end{aligned}$$

for any $s > 0$. We also have

$$\begin{aligned} P_m &= \Pr(\text{miss}) && (11)\\ &= \int \mathcal{I}(T(x) < \tau) f(x|\mathcal{K}) dx && (12)\\ &= \mathcal{E}\{\mathcal{I}(T(x) < \tau)|\mathcal{K}\} && (13)\\ &\leq \mathcal{E}\{e^{t(T(x)-\tau)}|\mathcal{K}\} && (14)\\ &= e^{-t\tau}\mathcal{E}\{e^{T(x)}|\mathcal{K}\} && (15) \end{aligned}$$

for any $t < 0$. Inequalities (10) and (15) are the most basic forms of the Chernoff bound. It is also easy to see that if $T(x)$ is linear in form, meaning

$$T(x) = \sum_{i=1}^{n} h_i(x_i) \tag{16}$$

we have

$$\begin{aligned} \mathcal{E}\{e^{sT(x)}|\mathcal{H}\} &= \mathcal{E}\{e^{s\sum_{i=1}^{n} h_i(x_i)}|\mathcal{H}\} && (17)\\ &= \prod_{i=1}^{n} \mathcal{E}\{e^{sh_i(x_i)}|\mathcal{H}\} && (18) \end{aligned}$$

which bound is (loosely-speaking) exponential in n. More precisely, if $h_i(\cdot) = h(\cdot)$ and the samples are identically distributed, then we can write

$$\mathcal{E}\{e^{T(x)}|\mathcal{H}\} = e^{n\log(\mathcal{E}\{e^{sh(x)}|\mathcal{H}\})} \tag{19}$$

Clearly (18) and (19) are true with $\mathcal{H}$ replaced by $\mathcal{K}$ and s replaced by t, as in (15). Note that all these are true for any $s > 0$ and $t < 0$. It hence behooves us[1] to find the "optimized" s and t that make the bounds tightest.

[1] . . . or else we should be hooved . . .

Things become even more interesting when we assume that $T(x)$ is the log-likelihood ratio ($T(x) = \log(L(x))$), and especially so when the data is conditionally *iid*. Define

$$\mu_{T,H}(s) \equiv \log(\mathcal{E}\{e^{sL(x)}|\mathcal{H}\}) \tag{20}$$
$$\mu_{T,K}(t) \equiv \log(\mathcal{E}\{e^{tL(x)}|\mathcal{K}\}) \tag{21}$$

so we have

$$P_f = e^{\mu_{T,H}(s)-s\tau} \tag{22}$$
$$P_m = e^{\mu_{T,K}(t)-t\tau} \tag{23}$$
$$\tag{24}$$

Now we can write

$$\begin{aligned}
\mu_{T,K}(t) &= \log(\mathcal{E}\{e^{t\log(L(x))}|\mathcal{K}\}) && (25)\\
&= \log(\mathcal{E}\{L(x)^t|\mathcal{K}\}) && (26)\\
&= \log(\int L(x)^t f(x|\mathcal{K})dx) && (27)\\
&= \log\left(\int \left(\frac{f(x|\mathcal{K})}{f(x|\mathcal{H})}\right)^t f(x|\mathcal{K})dx\right) && (28)\\
&= \log\left(\int \left(\frac{f(x|\mathcal{K})}{f(x|\mathcal{H})}\right)^{t+1} f(x|\mathcal{H})dx\right) && (29)\\
&= \mu_{T,H}(t+1) && (30)
\end{aligned}$$

Then we have

$$\begin{aligned}
P_m &= e^{\mu_{T,K}(t)-t\tau} && (31)\\
&= e^{\mu_{T,H}(t+1)-t\tau} && (32)\\
&= e^{\mu_{T,H}(s)-(s-1)\tau} && (33)\\
&= e^{\tau}e^{\mu_{T,H}(s)-s\tau} && (34)
\end{aligned}$$

meaning that both false alarm and miss bounds have the same form.

Now, since our search is for optimized bounds under $s > 0$ and $t < 0$, if we can minimize $\mu_{T,H}(s)$ for $0 < s < 1$ then we do not need to consider the two bounds separately at all. Hence let us define

$$\psi(s) = e^{\mu_{T,H}(s)-s\tau} \tag{35}$$
$$= \mathcal{E}\{e^{s(T(x)-\tau)}|\mathcal{H}) \tag{36}$$

Clearly we have

$$\begin{aligned}
\psi(0) &= 1 && (37)\\
\dot{\psi}(0) &= \mathcal{E}\{(T(x)-\tau)e^{s(T(x)-\tau)}|\mathcal{H})\Big|_{s=0} && (38)\\
&= \mathcal{E}\{T(x)|\mathcal{H}\}-\tau && (39)\\
\ddot{\psi}(0) &= \mathcal{E}\{(T(x)-\tau)^2e^{s(T(x)-\tau)}|\mathcal{H})\Big|_{s=0} && (40)\\
&\geq 0 && (41)
\end{aligned}$$

If we particularize to $T(x) = \log(L(x))$ we can go further, and write

$$\begin{aligned}
\dot{\psi}(1) &= \mathcal{E}\{(T(x)-\tau)e^{s(T(x)-\tau)}|\mathcal{H})\Big|_{s=1} && (42)\\
&= e^{-\tau}\int(\log(L(x))-\tau)e^{\log(L(x))}f(x|\mathcal{H})dx && (43)\\
&= e^{-\tau}\int(\log(L(x))-\tau)L(x)f(x|\mathcal{H})dx && (44)\\
&= e^{-\tau}\int(\log(L(x))-\tau)f(x|\mathcal{K})dx && (45)\\
&= \mathcal{E}\{T(x)|\mathcal{K}\}-\tau && (46)
\end{aligned}$$

It is reasonable to expect that

$$\begin{aligned}
\mathcal{E}\{\log(L(x))|\mathcal{H}\} &< \tau && (47)\\
\mathcal{E}\{\log(L(x))|\mathcal{K}\} &> \tau && (48)
\end{aligned}$$

Assuming so, we have from (40) that $\psi(s)$ is convex. Equations (37), (39) and (47) tell us that $\psi(s)$ passes through the point $(0,1)$ with a negative slope; and (46) and (48) that $\psi(s)$ has a positive slope when $s=1$. These facts together tell us that a minimum value for $\psi(s)$ (and therefore for $\mu_{T,H}(s)$) can be found for $0<s<1$, and hence we can combine the bounds (in this log-likelihood ratio case) as

$$\begin{aligned}
P_f &\leq e^{\mu_{T,H}(s_0)-s_0\tau} && (49)\\
P_m &\leq e^{\mu_{T,H}(s_0)+(1-s_0)\tau} && (50)
\end{aligned}$$

In the case that we have conditionally independent observations

$$\begin{aligned}
\mu(s) &= n\log\left(\int\left(\frac{f(x|\mathcal{K})}{f(x|\mathcal{H})}\right)^s f(x|\mathcal{H})dx\right) && (51)\\
&= n\log\left(\int f(x|\mathcal{K})^s f(x|\mathcal{H})^{1-s}dx\right) && (52)
\end{aligned}$$

and

$$\frac{d}{ds}\left(\mu(s) - s\tau\right)\Big|_{s=s_0} \;=\; 0 \tag{53}$$

for some $0 < s_0 < 1$. This specific form of Chernoff bound applies only to a log-likelihood ratio test, and (47) and (48) must hold. Note also that for *iid* data, $\mu(s) \propto n$ – this bound is exponential in the number of samples n.

3 The Bhattacharyya Bound

Note that (49) and (50) hold irrespective of s_0 as long as $0 < s_0 < 1$; we do not need the "optimal" s_0 as in (53). One good choice is $s_0 = 0.5$, for which

$$\mu(1/2) \;=\; \int \sqrt{f(x|\mathcal{H})f(x|\mathcal{K})}dx \tag{54}$$

This yields the Bhattacharyya bound, and it often both quite good and easy to use.

For the case that we are performing locally-optimal detection of a constant signal

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \nu_i \;+\; \theta \end{aligned} \tag{55}$$

in which $\{\nu_i\}$ are *iid* with pdf $f(x)$ and we additionally have $f(-x) = f(x)$ (a symmetric density, not so very uncommon) we can write for a single sample (to be multiplied by n in the bound)

$$\begin{aligned} e^{\mu_{T,H}(s)} &= n\log\left(\int f(x|\mathcal{K})^s f(x|\mathcal{H})^{1-s}dx\right) && (56)\\ &= n\log\left(\int f(x-\theta)^s f(x)^{1-s}dx\right) && (57)\\ &= n\mu(s) && (58)\\ e^{\mu_{T,H}(1-s)} &= n\log\left(\int f(x-\theta)^{1-s} f(x)^{s}dx\right) && (59)\\ &= n\log\left(\int f(\theta-x)^{1-s} f(x)^{s}dx\right) && (60)\\ &= n\log\left(\int f(y)^{1-s} f(\theta-y)^{s}dx\right) && (61)\\ &= n\log\left(\int f(y)^{1-s} f(y-\theta)^{s}dx\right) && (62)\\ &= n\mu(s) && (63) \end{aligned}$$

which implies that the bound is minimized by $s = 1/2$ – the Chernoff bound is the Bhattacharyya bound in such cases. But the Bhattacharyya bound is always valid; it is not always the tightest bound, but it is easy to find and to use.

4 Examples

4.1 Laplace Noise

Suppose we have

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \nu_i \;+\; \theta \end{aligned} \tag{64}$$

with threshold $\tau = 0$, log-likelihood ratio detection and $\{\nu_i\}$ *iid* Laplace

$$f(x) \;=\; \frac{1}{2} e^{-|x|} \tag{65}$$

Since $f(x)$ is symmetric we know the applicable Chernoff bound is that with $s = 1/2$ – the Bhattacharyya bound. We write

$$\begin{aligned} \mu(1/2) &= n \log \left(\int_{-\infty}^{\infty} \sqrt{\frac{1}{2} e^{-|x|} \frac{1}{2} e^{-|x-\theta|}} dx \right) && (66) \\ &= n \log \left(\int_{-\infty}^{0} \frac{1}{2} e^{x-\theta/2} dx \;+\; \int_{0}^{\theta} \frac{1}{2} e^{-\theta/2} dx \right. \\ &\qquad \left. + \int_{\theta}^{\infty} \frac{1}{2} e^{-x+\theta/2} dx \right) && (67) \\ &= n \log \left(\frac{1}{2} e^{-\theta/2} \;+\; \frac{\theta}{2} e^{-\theta/2} \;+\; \frac{1}{2} e^{-\theta/2} \right) && (68) \\ &= n \log \left((1 + \theta/2) e^{-\theta/2} \right) && (69) \end{aligned}$$

For the case that $\theta = 1$ (note that we do not need to make a small-signal assumption with Chernoff) we get

$$\mu(1/2) \;\approx\; -0.0945 n \tag{70}$$

$$\begin{aligned} P_f &\leq e^{-0.0945n} && (71) \\ P_m &\leq e^{-0.0945n} && (72) \end{aligned}$$

since the threshold is zero.

4.2 Exponentials

Suppose we have

$$\mathcal{H}: \quad x_i \sim e^{-x}u(x) \tag{73}$$
$$\mathcal{K}: \quad x_i \sim 2e^{-2x}u(x) \tag{74}$$

and again[2] $\tau = 0$. We have

$$\begin{aligned}
\mu(s) &= n\log\left(\int (2e^{-2x})^s (e^{-x})^{1-s} dx\right) && (75)\\
&= n\log\left(\int_0^\infty 2^s e^{-sx} e^{-x} dx\right) && (76)\\
&= n\log\left(\frac{2^s}{s+1}\right) && (77)\\
&= n(s\log(2) - \log(s+1)) && (78)
\end{aligned}$$

We differentiate to find

$$\begin{aligned}
0 &= \frac{d}{ds}\mu(s) && (79)\\
&= \log(2) - \frac{1}{s+1} && (80)
\end{aligned}$$

hence

$$s_0 = \frac{1}{\log(2)} - 1 \approx 0.443 \tag{81}$$

We substitute to find

$$\mu(s_0) = n\left(1 - \log(2) + \log(\log(2))\right) \approx -0.0597n \tag{82}$$

hence

$$\begin{aligned}
P_f &\leq e^{-0.0597n} && (83)\\
P_m &\leq e^{-0.0597n} && (84)
\end{aligned}$$

since the threshold is zero.

4.3 Exponentials with Threshold

Let us use the same example as (74), except that the testing threshold (for the log-likelihood ratio) is $\tau \neq 0$. We still have (78), but now we must

[2]Please be aware that with nonzero τ the full (53) must be solved.

optimize by differentiating $\mu(s) - s\tau$. We get

$$\begin{aligned} 0 &= \frac{d}{ds}(\mu(s) - s\tau) && (85) \\ &= n\left(\log(2) - \frac{1}{s+1}\right) - \tau && (86) \end{aligned}$$

hence

$$s_0 = \frac{1}{\log(2e^{-\tau/n})} - 1 \quad (87)$$

Now suppose we choose $\tau = 2n$. We get $s_0 = -1.765$, and clearly this does not make sense in that we know $0 < s_0 < 1$. The reason is illuminating. For the test (74) we know that the LLR is

$$T(x) = \log(2) - x \quad (88)$$

meaning that its maximum (single sample) value is $\log(2) \approx 0.693$.

That is, this test amounts to $\delta(x) = 0$ if $\tau > \log(2)$. So let us try a non-trivial threshold: $\tau = 0.5n$. We get $s_0 = 4.177$ and hence $\mu = 1.251$ which implies

$$\begin{aligned} P_f &\leq e^{\mu(s_0) - s_0\tau} && (89) \\ &= e^{-0.833n} && (90) \end{aligned}$$

Note that this does not work for the miss probability, since $s_0 > 1$. Why is this? The reason is (47) and (48), repeated

$$\mathcal{E}\{\log(L(x))|\mathcal{H}\} < \tau < \mathcal{E}\{\log(L(x))|\mathcal{K}\} \quad (91)$$

which was the requirement for the minimizing $0 < s_0 < 1$. Now since

$$\begin{aligned} \mathcal{E}\{\log(L(x))|\mathcal{H}\} &= \mathcal{E}\{\log(2) - x)|\mathcal{H}\} = \log(2) - 1 = -0.307 && (92) \\ \mathcal{E}\{\log(L(x))|\mathcal{K}\} &= \mathcal{E}\{\log(2) - x)|\mathcal{H}\} = \log(2) - 1/2 = 0.193 && (93) \end{aligned}$$

we see that for a valid Chernoff bound for both P_f and P_m is possible only if $-0.3069n < \tau < 0.1931$. In fact, returning to (33) we see its minimizing nonpositive $s = 1$, hence the best Chernoff can do is $P_m \leq 1$.

To be legal, let us choose $\tau = 0.1n$; we get

$$\begin{aligned} s_0 &= 0.686 && (94) \\ \mu(s_0) &= -0.047 && (95) \end{aligned}$$

and hence

$$\begin{aligned} P_f &\leq e^{\mu(s_0) - s_0\tau} = e^{-0.116n} && (96) \\ P_m &\leq e^{\mu(s_0) - (1-s_0)\tau} = e^{-0.078n} && (97) \end{aligned}$$

which, due to the non-zero threshold, are not the same.

5 Importance Sampling

5.1 Estimation of Small Probabilities

Suppose we want to estimate a small probability

$$\alpha \equiv Pr(x \in \Omega) \tag{98}$$

This may sound trivial, and it would be if we were interested, say, in the $\Omega = \{x : x > \tau\}$. But suppose it is not so simple, and Ω is the set of noise samples[3] that produces an error in an OFDM system with LDPC, zero-forcing equalization and carrier-offset recovery. We have no hope of an analytic probability calculation, all we can do is simulate and count the errors. That is, we estimate

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{I}(x_i \in \Omega) \tag{99}$$

where $\mathcal{I}$ is the indicator, N is the number of Monte Carlo trials, these indexed by i. It is vary simple to see that

$$\mathcal{E}\{\hat{\alpha}\} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}\{\mathcal{I}(x_i \in \Omega)\} \tag{100}$$

$$= \int_{\Omega} f(x)dx = \alpha \tag{101}$$

which is good news, but

$$\frac{Var\{\hat{\alpha}\}}{(\mathcal{E}\{\hat{\alpha}\})^2} \approx \frac{\alpha/N}{\alpha^2} = \frac{1}{N\alpha} \tag{102}$$

which is not good news. Equation (102) means that if you want the standard deviation of $\hat{\alpha}$ to be (say) less than 10% of its value, you need $N > 100/\alpha$ MC trials; and if α is 10^{-8} this can be a chore.

Fortunately we have *importance sampling*[4] to help. Consider a new estimator

$$\hat{\alpha} = \sum_{i=1}^{N}\mathcal{I}(x_i \in \Omega)\frac{f(x_i)}{q(x_i)} \tag{103}$$

[3] Don't worry if this OFDM stuff means nothing to you; the point is that it's a complicated event.

[4] Much of this section is taken from the notes for Advanced Signal Processing, ECE 6123.

where $f(\cdot)$ is the true probability density governing whatever is random about your problem and $q(\cdot)$ some other "importance" pdf that the samples used to estimate $\hat{\alpha}$ are actually drawn from. For example, we might have $\Omega = \{(x_1, x_2) : (x_1 - 10)^2 + (x_2 - 15)^2 \leq 1\}$ and $f(\cdot)$ bivariate Gaussian with mean zero and unity variance. It is fairly clear that $\hat{\alpha}$ from (99) will include exactly no indicators that "happen" for any reasonable value of N – it is useless. But suppose we use (103) with $q(\cdot)$ to mean a Gaussian pdf with mean $(10, 15)$ and variance of 0.5: then many indicators will fire, and each of them will force the inclusion to the sum in (103) of many relatively small values determined by the *importance weights* $p(x_i)/q(x_i)$.

The variance of (103) is easily seen to be

$$Var(\hat{\alpha}) \;=\; \frac{1}{N}\left(\int_\Omega \frac{f(x)^2}{q(x)}dx - \alpha^2\right) \tag{104}$$

which is illuminating for two reasons. The first is that it is minimized (actually cut down to zero) by

$$q(x) \;=\; f(x|x \in \Omega) \;=\; \frac{f(x)\mathcal{I}(x \in \Omega)}{\alpha} \tag{105}$$

which gives us the helpful information that if we already knew the answer we could easily use MC techniques to find the answer. This actually really *is* useful, since it tells us that there is no "magic bullet" for importance-sampling – choosing a good $q(\cdot)$ is an art form. But (104) also suggests intuition: if we want to have a low variance we should try to reduce the variation of $f(\cdot)/q(\cdot)$ of Ω as much as we can. By that logic choosing $q(\cdot)$ to have mean $(9, 14)$ and unity variance in the previous example may be better than the $q(\cdot)$ given; and mean mean $(11, 16)$ worse.

It is worth mentioning that in the case of binary hypothesis testing, a good importance density is

$$q(x) \;=\; f(x|\mathcal{K}) \tag{106}$$

when used to estimate the probability of false alarm α. Since generally P_d (i.e., β) is usually not that extreme, importance sampling is likely not necessary under $\mathcal{K}$.

5.2 Importance Sampling for Moments

Consider the situation as in figure 1, in which we wish to calculate the expected value of a function $g(x)$ under the pdf $f(x)$. A direct MC implementation will probably not work very well, since the "active" part of

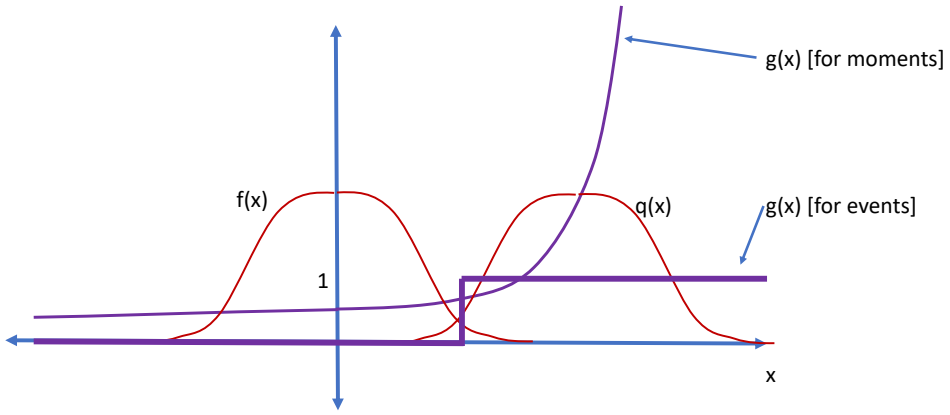

Figure 1: Illustration of the way the true pdf ($f(x)$), the importance density ($q(x)$) and the desired function whose expected value is being sought ($g(x)$).

$g(x)$ occurs where samples from $p(x)$ are rare. Suppose instead we simulate under $q(x)$ as also indicated in the plot. Then we get

$$\bar{g} \;=\; \frac{1}{N}\sum_{i=1}^{N} g(x_i)\frac{f(x_i)}{q(x_i)} \tag{107}$$

so that

$$\mathcal{E}\{\bar{g}(x)\} \;=\; \int g(x)\frac{f(x)}{q(x)}q(x)dx \;=\; \mathcal{E}\{g(x)\} \tag{108}$$

meaning that the importance-sampling estimator is unbiased for this case, too.

ECE 6124
Signal Detection
Sequential and Quickest Detection

Peter Willett

Fall 2018

1 Introduction

There are two main sections in this chapter: sequential detection and quickest detection. They are both related by their sequential "anytime" nature and their mathematical underpinnings. However, they differ strongly in the problems that they attack. Sequential detection refers to a rather simpler problem, more closely related to the standard hypothesis testing we have previously seen; quickest detection is rather different.

In a sequential detection problem either $\mathcal{H}$ or $\mathcal{K}$ is true. Our aim is to save time: if the evidence fo $\mathcal{K}$ (say) mounts to an almost overwhelming certainty before the detection record of $1 \leq i \leq n$ is up – maybe at $i = n/2$, then why continue? Why not save time by making the decision now? To make it concrete, suppose we are interested in

$$\begin{aligned} \mathcal{H}: \quad & x_i = \nu_i - 1 \\ \mathcal{K}: \quad & x_i = \nu_i + 1 \end{aligned} \tag{1}$$

and we use

$$T(x) = \sum_{i=1}^{10} x_i \tag{2}$$

and $\tau = 5$. Now suppose after 5 samples we have $T(x) = 8$; that means

$$Pr(\delta = 0|\mathcal{H}) = 0.10 \tag{3}$$

$$Pr(\delta = 0|\mathcal{K}) = 0.005 \tag{4}$$

so perhaps simply ending the test makes sense. The key challenges for sequential detection are to develop the optimal scheme – the *Wald test* or SPRT – to understand what optimal means in this context, and compute the performance, which will turn out to be in terms of run-length as opposed to α and β.

In a quickest detection problem $\mathcal{H}$ is always true at the start. There may be a switch to $\mathcal{K}$ at some point, in which case it would be desirable to

know it as quickly as possible. On the other hand there may be no change ($\mathcal{H}$ always remains true). The notional problem here is the assembly line: let's suppose that under nominal conditions the factory produces gearboxes that are non-functional 1% of the time; when the line is malfunctioning this imperfection rate increases to 5%. Note that bad gearboxes happen in both situations, and the key is to figure when enough bad gearboxes have recently appeared that one can raise an alarm. This problem does not at first blush appear related to our usual detection problem nor to the sequential problem. However, the SPRT is the basis for the optimal *Page procedure* or CUSUM, and the issues that arise are to define optimality and to determine the run-lengths: the average times between false alarms and the average delay to detection. As with the sequential problem the probabilities of error are of less interest, and in fact any nontrivial procedure must have any reasonably-defined $\alpha = \beta = 1$, since eventually an alarm will be raised no matter the situation. The issue is: how often?

We will also show, briefly, that when the change-of-hypothesis process that underlies the observations can be modeled as Markov – that is, the observations process is a hidden Markov model (HMM) – then there is an appropriate Bayesian approach that attempts to estimate the underlying Markov state. This is called the *Shiryaev* testing procedure.

Note that both sequential detection and quickest detection have their place in target tracking. Specifically, once a tentative track is initiated a *sequential test* begins: if it passes then the track becomes confirmed; if it fails, the tentative track is deleted. But once a track is confirmed a *quickest detection* procedure should immediately begin, with null hypothesis $\mathcal{H}$ that the track is extant; once it declares for the alternative $\mathcal{K}$ that the track no longer exists, the track is terminated.

2 Sequential Detection

2.1 The SPRT

With the usual definitions of our simple test – $f(x|\theta)$, $\Theta_H = \theta_H$ and $\Theta_K = \theta_K$ – we additionally define the stopping rule $\phi_n(x) \in \{0, 1\}$ such that

$$N_\phi \equiv \{\phi_N(x) = 1 \text{ and } \phi_m(x) = 0 \; \forall \; 1 \leq m < N\} \tag{5}$$

and a sequence of decision rules $\{\delta_n(x)\}$ such that the final decision is $\delta_N(x)$.

The sequential probability ratio test (SPRT or Wald test) uses

$$\phi_n(x) = \begin{cases} 1 & L_n(x) \geq B \\ 1 & L_n(x) \leq A \\ 0 & A < L_n(x) < B \end{cases} \tag{6}$$

$$\delta_n(x) = \begin{cases} 1 & L_n(x) \geq B \\ 0 & L_n(x) \leq A \\ - & A < L_n(x) < B \end{cases} \tag{7}$$

in which $L_n(x)$ is the likelihood ratio of the first n samples of x. The appearance of the likelihood ratio is not surprising. The last line in (7) indicates a "don't call" situation.

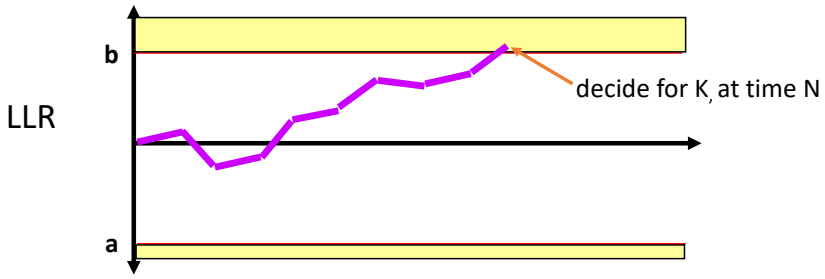

Figure 1: Notional figure of an SPRT. In this case we are using the logarithm of the likelihood ratio. The impact of this on the designed thresholds will be discussed shortly.

The means to set the thresholds A and B may indeed be surprising, however. Suppose we write

$$Q_n^H \equiv \{x : N_\phi = n \text{ and } \delta_n(x) = 0\} \tag{8}$$

$$Q_n^K \equiv \{x : N_\phi = n \text{ and } \delta_n(x) = 1\} \tag{9}$$

and note that for $n \neq m$ we know

$$Q_n^H \bigcap Q_m^H = \emptyset \tag{10}$$

$$Q_n^K \bigcap Q_m^K = \emptyset \tag{11}$$

It is clear that we can define

$$\Omega_H \equiv \bigcup_{n=1}^{\infty} Q_n^H \tag{12}$$

$$\Omega_K \equiv \bigcup_{n=1}^{\infty} Q_n^K \tag{13}$$

as the decision region for $\mathcal{K}$.

Now we have

$$\begin{aligned}
\alpha &= Pr(x \in \Omega_K | \mathcal{H}) && (14)\\
&= \sum_{n=1}^{\infty} Pr(x \in Q_n^K | \mathcal{H}) && (15)\\
&= \sum_{n=1}^{\infty} \int_{Q_n^K} f(x|\theta_H) && (16)\\
&\leq \sum_{n=1}^{\infty} \int_{Q_n^K} \frac{1}{B} f(x|\theta_K) && (17)\\
&= \frac{1}{B} \sum_{n=1}^{\infty} Pr(x \in Q_n^K | \mathcal{K}) && (18)\\
&= \frac{\beta}{B} && (19)
\end{aligned}$$

in which (17) follows because for Q_n^K we must have $L_n(x) \geq B$. We also have

$$\begin{aligned}
1-\beta &= Pr(x \in \Omega_H | \mathcal{K}) && (20)\\
&= \sum_{n=1}^{\infty} Pr(x \in Q_n^H | \mathcal{K}) && (21)\\
&= \sum_{n=1}^{\infty} \int_{Q_n^H} f(x|\theta_K) && (22)\\
&\leq \sum_{n=1}^{\infty} \int_{Q_n^H} A f(x|\theta_H) && (23)\\
&= A \sum_{n=1}^{\infty} Pr(x \in Q_n^H | \mathcal{H}) && (24)\\
&= A(1-\alpha) && (25)
\end{aligned}$$

in which (23) follows because for Q_n^H we must have $L_n(x) \leq A$. Putting (19) together with (25) we have

$$\begin{aligned}
\alpha &\leq \frac{\beta}{B} && (26)\\
1-\beta &\leq A(1-\alpha) && (27)
\end{aligned}$$

Suppose we treat these as equality, under the reasonable assumption that the test statistic barely scrapes above or below the corresponding threshold

– the "excess over the boundaries" is negligible. We get

$$A = \frac{1-\beta}{1-\alpha} \tag{28}$$

$$B = \frac{\beta}{\alpha} \tag{29}$$

in which it musts be remembered that these are thresholds that the *likelihood ratio* should use – if one uses the $T(x) = \log(L(x))$ one should use $\log(A)$ and $\log(B)$, and of course if some other transformation $g(L(x))$ is to be used then $g(\cdot)$ should be applied to the thresholds.

It is quite remarkable that the thresholds are given explicitly in terms of the desired false alarm and missed detection rates – in this regard, then, the sequential situation is simpler than fixed-length detection. Note also that if we use these desired α_d and β_d to design the thresholds, the actual performance, from (19) and (25), is

$$\alpha \leq \frac{\beta}{B} \tag{30}$$

$$\leq \frac{\alpha_d \beta}{\beta_d} \tag{31}$$

$$\leq \frac{\alpha_d}{\beta_d} \tag{32}$$

$$1-\beta \leq A(1-\alpha) \tag{33}$$

$$\leq \frac{1-\beta_d}{1-\alpha_d}(1-\alpha) \tag{34}$$

$$\leq \frac{1-\beta_d}{1-\alpha_d} \tag{35}$$

so assume both α_d and $1-\beta_d$ there is little loss.

2.2 The Expected Run Lengths

First, note that this formulation applies to *iid* samples. Suppose we define the running sum as

$$s_n = \sum_{i=1}^{n} \log\left(\frac{f(x_i|\mathcal{K})}{f(x_i|\mathcal{H})}\right) \tag{36}$$

$$= \sum_{i=1}^{n} y_i \tag{37}$$

which applies to the log-likelihood ratio – this will be our assumed test format. Let us define

$$p_{nj}(s_n) \equiv \text{the pdf of } s_n \text{ given } j \text{ is true} \tag{38}$$

$$
\begin{aligned}
p_{nj}(s_n|N=m) &\equiv \text{the pdf of } s_n \text{ given } j \text{ is true and } N_\phi = m \leq n \quad (39)\\
p_{nj}(s_n|N\leq n) &\equiv \text{the pdf of } s_n \text{ given } j \text{ is true and } N_\phi \leq n \quad (40)\\
q_{nj}(s_n) &\equiv \text{the pdf of } s_n \text{ given } j \text{ is true and } N_\phi = n \quad (41)
\end{aligned}
$$

in which $j \in \{\mathcal{H}, \mathcal{K}\}$. Now it is easy to see that we have

$$
p_{nj}(s_n|N=m) \;=\; \int q_{mj}(\sigma) p_{(n-m)j}(s_n - \sigma) d\sigma \quad (42)
$$

Now we can also have

$$
p_{nj}(s_n|N\leq n) \;=\; \sum_{m=1}^{n} \int q_{mj}(\sigma) p_{(n-m)j}(s_n - \sigma) d\sigma \frac{Pr(N_\phi = m|j)}{Pr(N_\phi \leq m|j)} \quad (43)
$$

Let us assume that n is large so that for most of the terms in the sum $Pr(N_\phi \leq m|j) \approx 1$:

$$
p_{nj}(s_n|N\leq n) \;=\; \sum_{m=1}^{n} \int q_{mj}(\sigma) p_{(n-m)j}(s_n - \sigma) d\sigma Pr(N_\phi = m|j) \quad (44)
$$

Let us take the moment generating functions

$$
\begin{aligned}
\psi_{mj}(t) &\equiv \mathcal{E}\{e^{ts_m}|j\} \quad (45)\\
&= \int e^{ts_m} q_{mj}(s_m) ds_m \quad (46)
\end{aligned}
$$

and

$$
M_j(t) \;\equiv\; \mathcal{E}\{e^{ty_i}|j\} \quad (47)
$$

for just a single sample. Clearly we have for a fixed n

$$
\begin{aligned}
\mathcal{E}\{e^{ts_n}\} &= \int e^{ts_n} p_{nj}(s) ds_n \quad (48)\\
&= [M_j(t)]^n \quad (49)
\end{aligned}
$$

Examining (44) and noting that n is very large – and hence a decision must have been made prior to n: $N_\phi \leq n$ – we take the moment-generating function of both sides of (44). The left-hand side is the same as (49); the same is true for the p term on the right. We have

$$
[M_j(t)]^n \;=\; \sum_{m=1}^{n} \psi_{mj}(t) \, [M_j(t)]^{n-m} \, Pr(N_\phi = m|j) \quad (50)
$$

We simplify and come up with the remarkable

$$\begin{aligned} 1 &= \sum_{m=1}^{n} \psi_{mj}(t) \left[M_j(t)\right]^{-m} Pr(N_\phi = m|j) && (51)\\ &= \sum_{m=1}^{n} \mathcal{E}\{e^{ts_m}|j\} \left[M_j(t)\right]^{-m} Pr(N_\phi = m|j) && (52)\\ &= \mathcal{E}\{e^{ts_N} \left[M_j(t)\right]^{-N} |j\} && (53) \end{aligned}$$

where we have, for the sake of pulchritude, used N to denote N_ϕ. The marvelous last relation, (53), is commonly known as *Wald's Identity.*

Now, let us differentiate (53):

$$\begin{aligned} &\frac{d}{dt}\left(\mathcal{E}\{e^{ts_N} \left[M_j(t)\right]^{-N} |j\}\right)\\ &= \mathcal{E}\{s_N e^{ts_N} \left[M_j(t)\right]^{-N} - N\dot{M}_j(t) \left[M_j(t)\right]^{-N-1} |j\} && (54) \end{aligned}$$

Now by definition, we have

$$\begin{aligned} M_j(t) &\equiv \mathcal{E}\{e^{ty_i}|j\} && (55)\\ M_j(0) &= 1 && (56)\\ \dot{M}_j(0) &= \mathcal{E}\{y_i|j\} && (57)\\ \ddot{M}_j(0) &= \mathcal{E}\{y_i^2|j\} && (58) \end{aligned}$$

Hence we evaluate (54) at $t = 0$ to get

$$\mathcal{E}\{s_N|j\} = \mathcal{E}\{N|j\}\mathcal{E}\{y_i|j\} \quad (59)$$

Note that we know $\mathcal{E}\{y_i|j\}$ trivially; and $\mathcal{E}\{s_N|j\}$ can be approximated straightforwardly from the thresholds and the error probability. In the case that $\mathcal{E}\{y_i|j\} = 0$, meaning that (59) provides us no information about $\mathcal{E}\{N|j\}$, we differentiate (54) again to get

$$\begin{aligned} &\frac{d^2}{dt^2}\left(\mathcal{E}\{e^{ts_N} \left[M_j(t)\right]^{-N} |j\}\right)\\ &= \frac{d}{dt}\mathcal{E}\left\{\left[s_N - N\left(\frac{\dot{M}_j(t)}{M_j(t)}\right)\right] e^{ts_N} \left[M_j(t)\right]^{-N} \middle| j\right\} && (60)\\ &= \mathcal{E}\left\{\left(\left[s_N - N\left(\frac{\dot{M}_j(t)}{M_j(t)}\right)\right]^2 - N\frac{\ddot{M}_j(t)M_j(t) - \dot{M}_j(t)^2}{M_j(t)^2}\right)\right.\\ &\qquad \left. \times e^{ts_N} \left[M_j(t)\right]^{-N} \middle| j\right\} && (61) \end{aligned}$$

We evaluate (61) at $t = 0$ and, assuming $\dot{M}(t) = 0$, get

$$\mathcal{E}\{s_N^2|j\} \;=\; \mathcal{E}\{N|j\}\mathcal{E}\{y_i^2|j\} \tag{62}$$

The second moments can be read as variances since $\mathcal{E}\{y_i|j\} = 0$. Note that a log-likelihood ratio will not have a zero mean under either hypothesis. However, Wald's Identity holds for more than just optimal tests.

As an example, consider

$$\begin{aligned} \mathcal{H}: \quad & x_i \;=\; \nu_i \\ \mathcal{K}: \quad & x_i \;=\; \nu_i \;+\; 1 \end{aligned} \tag{63}$$

in which $\{\nu_i\}$ are unit normal and independent. We design for $\alpha = 0.1\%$ and $\beta = 99\%$. Via Wald's Approximations we have

$$B \;=\; \frac{\beta}{\alpha} \;=\; \frac{.99}{.001} \;=\; 990 \tag{64}$$

$$b \;=\; \log(B) \;=\; 6.9 \tag{65}$$

$$A \;=\; \frac{1-\beta}{1-\alpha} \;=\; \frac{.01}{.999} \;=\; 0.01001 \tag{66}$$

$$a \;=\; \log(B) \;=\; -4.6 \tag{67}$$

We also have

$$y_i \;=\; \log\left(\frac{f(x_i)|\mathcal{K}}{f(x_i)|\mathcal{H}}\right) \;=\; x_i - 0.5 \tag{68}$$

Now

$$\mathcal{E}\{y_i|\mathcal{H}\} \;=\; -0.5 \tag{69}$$

$$\mathcal{E}\{y_i|\mathcal{K}\} \;=\; +0.5 \tag{70}$$

$$\mathcal{E}\{s_N|\mathcal{H}\} \;=\; (1-\alpha)a + \alpha b \;=\; -4.59 \tag{71}$$

$$\mathcal{E}\{s_N|\mathcal{K}\} \;=\; (1-\beta)a + \beta b \;=\; 6.79 \tag{72}$$

Hence

$$\mathcal{E}\{N|\mathcal{H}\} \;=\; \frac{\mathcal{E}\{s_N|\mathcal{H}\}}{\mathcal{E}\{y_i|\mathcal{H}\}} \;=\; 9.2 \tag{73}$$

$$\mathcal{E}\{N|\mathcal{K}\} \;=\; \frac{\mathcal{E}\{s_N|\mathcal{K}\}}{\mathcal{E}\{y_i|\mathcal{K}\}} \;=\; 13.6 \tag{74}$$

It is interesting to compare this to a fixed length test that achieves the same performance. We have

$$\alpha \;=\; Q\left(\frac{\tau - N\mathcal{E}\{y_i|\mathcal{H}\}}{\sqrt{N\mathcal{V}\{y_i|\mathcal{H}\}}}\right) \tag{75}$$

$$\tau \;=\; \sqrt{N}Q^{-1}(.001) - N/2 \tag{76}$$

hence

$$\beta = Q\left(\frac{\tau - N\mathcal{E}\{y_i|\mathcal{K}\}}{\sqrt{N\mathcal{V}\{y_i|\mathcal{K}\}}}\right) \quad (77)$$

$$0.99 = Q\left(Q^{-1}(.001) - \sqrt{N}\right) \quad (78)$$

This can be solved to get $N \approx 30$. It is actually typical to get a gain of 2-3 via sequential testing.

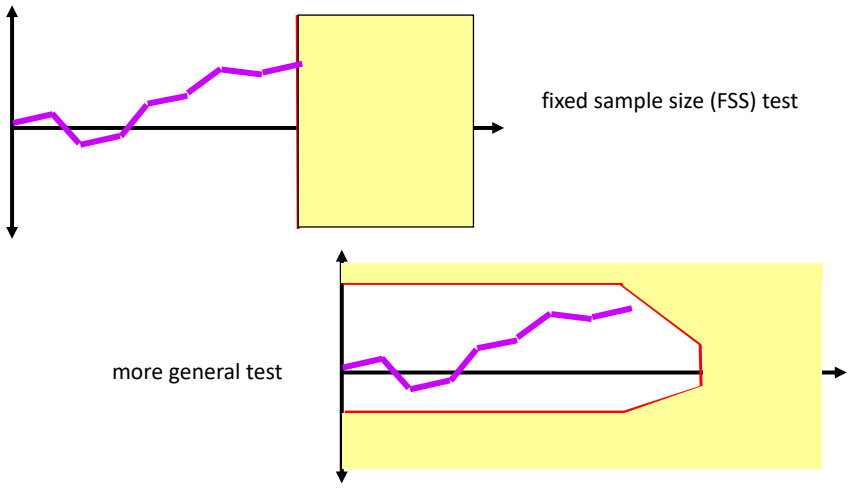

Figure 2: These are two different tests. The upper cartoon illustrates the fixed-length test, and the lower a more free-wheeling test that has fixed-length and sequential characteristics.

2.3 The Wald-Wolfowitz Theorem

The notion of optimality is rather slippery in the sequential situation: we could always improve α and β if we took more samples; and we could always get by with fewer samples if we absorbed the hit in α and β. The Wald-Wolfowitz theorem formalizes this.

Consider the SPRT (ϕ^w, δ^w) and any other sequential procedure (ϕ, δ). The theorem states that if

$$P_f(\phi, \delta) \leq P_f(\phi^w, \delta^w) \quad (79)$$

$$P_d(\phi, \delta) \geq P_d(\phi^w, \delta^w) \quad (80)$$

then we must have

$$\mathcal{E}\{N_\phi|\mathcal{H}\} \geq \mathcal{E}\{N_{\phi^w}|\mathcal{H}\} \tag{81}$$
$$\mathcal{E}\{N_\phi|\mathcal{K}\} \geq \mathcal{E}\{N_{\phi^w}|\mathcal{K}\} \tag{82}$$

In words then, any test that causes fewer errors than the SPRT must take longer to do so. With reference to figure 2 this is perhaps surprising.

Consider that we formulate a stopping rule that is similar to the lower plot: that is, we use the SPRT, but decide that after a certain number of samples – say, the expected number of samples the SPRT takes – we simply stop and make a decision. Clearly the expected number of samples such a strategy would use is no more than the SPRT with those upper and lower thresholds would use. However, the Wald-Wolfowitz Theorem insists that this does not apply, since the probabilities of error must exceed the SPRT's. The Theorem goes further, however: it says that if an attempt were made to adjust the upper and lower thresholds outwards so that the error probabilities matched the SPRT's, that attempt would end in failure since the thresholds would have to be infinite. Alternatively, if the "hard decision" boundary were placed somewhere greater than the SPRT's expected run length, then the upper and lower thresholds *could* be chosen to match the SPRT's error probabilities; but then the expected run length of this modified test would, disappointingly, exceed that of the SPRT. It is surprising.

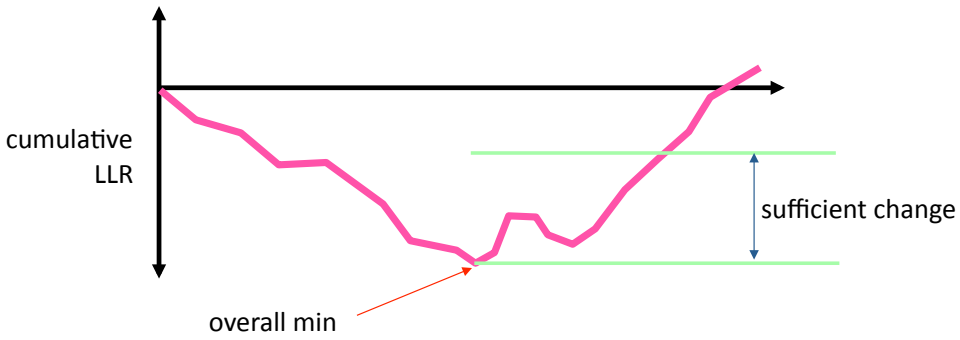

Figure 3: The notional sequential procedure to alert for a change in distribution. The idea, from (84), is to see if the recent evidence for $\mathcal{K}$ is sufficient.

3 Quickest Detection

3.1 The Page Test

Here we are not interested in the traditional detection problem in which either $\mathcal{H}$ or $\mathcal{K}$ is true. In quickest detection the samples always begin (at time $i = 0$) as governed by $\mathcal{H}$; and there may be a change to $\mathcal{K}$ at some unknown time n_0. That is,

$$\begin{aligned} x_i &\sim f(x_i|\mathcal{H}) \quad i = 1, 2, \ldots, n_0 \\ x_i &\sim f(x_i|\mathcal{K}) \quad i = n_0 + 1, n_0 + 2, \ldots \end{aligned} \tag{83}$$

and all are assumed independent. Intuition suggests the following *sequential* procedure:

$$\text{Declare a change has taken place when } \Sigma_n - \min_{m<n}\{\Sigma_m\} \; > \; h \tag{84}$$

in which

$$\Sigma_n \; \equiv \; \sum_{i=1}^{n} \log\left(\frac{f(x_i|\mathcal{K})}{f(x_i|\mathcal{H})}\right) \tag{85}$$

With reference to figure 3 we see that the procedure might be thought equivalent to the *Page* or *CUSUM* (cumulative sum[1]) procedure, which is

$$\text{Declare a change has taken place when } S_n \; > \; h \tag{86}$$

in which

$$S_n \; \equiv \; \max\left\{S_{n-1} + \log\left(\frac{f(x_n|\mathcal{K})}{f(x_n|\mathcal{H})}\right), 0\right\} \tag{87}$$

is the CUSUM. Actually any

$$S_n \; \equiv \; \max\left\{S_{n-1} + g(x_n), 0\right\} \tag{88}$$

is fine as long as

$$\mathcal{E}\{g(x_i)|\mathcal{H}\} \; < \; 0 \tag{89}$$

$$\mathcal{E}\{g(x_i)|\mathcal{K}\} \; > \; 0 \tag{90}$$

This is pictured in figure 4.

Optimality is only assured if $g(x)$ is the log-likelihood ratio; but since we want to be able to use the Page procedure in cases in which we are not sure of the hypotheses[2]

[1] The Matlab command "cumsum" does not implement the CUSUM.

[2] We probably know $f(x_n|\mathcal{H})$, either by model or because we can estimate it fairly well based on the recent history. Hence uncertainly is especially applicable to the case that $f(x_n|\mathcal{K})$ is not known precisely, meaning that we may not know how "strong" the change is. But any $g(\cdot)$ that satisfies (90) is acceptable.

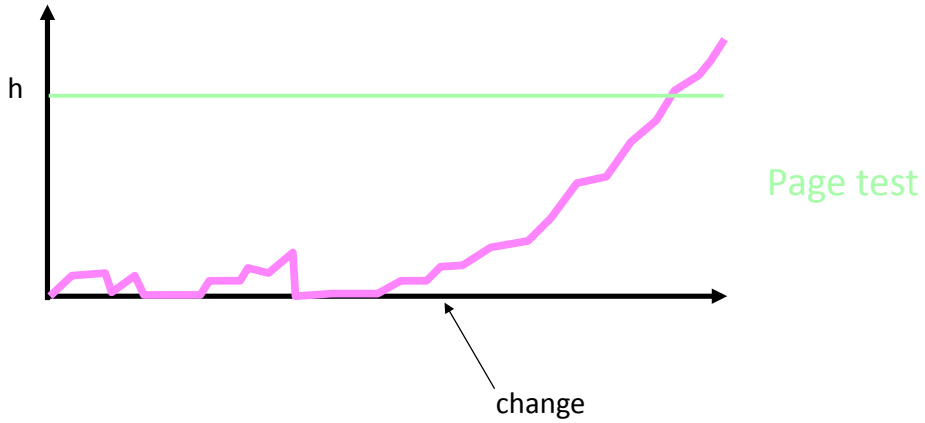

Figure 4: The Page procedure to alert for a change in distribution. The CUSUM, from (86), is equivalent to what is shown in figure 3.

The Page idea actually works remarkably well. In figure 5 we show a change from $\mathcal{N}(-0.2, 1)$ to $\mathcal{N}(+0.2, 1)$ at time $n = 100$. The change is not really perceptible to the eye, at least not from the data itself; however the Page procedure finds it easily, with an acceptable delay. See also figure 6.

Now, as indicated in the early discussion, α and β are not really at issue here: both are unity unless the situation is trivial. Hence performance is measured in terms of average delay to detection $\bar{D}$ and average time between false alarms $\bar{T}$. The Page procedure can be shown to be optimal for independent data[3] in these terms, in that for a given $\bar{T}$ any other test has at least as large a $\bar{D}$.

3.2 Performance

For now, let us model a sequential test with lower threshold b and upper threshold h: $b < 0 < h$ and we assume that $\mathcal{E}\{g(x_i)|\mathcal{H}\} < 0 < \mathcal{E}\{g(x_i)|\mathcal{K}\}$ as in (90). We define

$$\begin{aligned} P(\theta) &= Pr(\text{Wald test } (b,h) \text{ ends at } b) && (91)\\ M(\theta) &= \mathcal{E}\{\text{number of samples in Wald test } (b,h)\} && (92)\\ M_h(\theta) &= \mathcal{E}\{\text{number of samples in Wald test } (b,h) \text{ ending at } h\} && (93)\\ M_b(\theta) &= \mathcal{E}\{\text{number of samples in Wald test } (b,h) \text{ ending at } b\} && (94) \end{aligned}$$

[3]For dependent data there are results – for example of Markov processes – but little of general applicability.

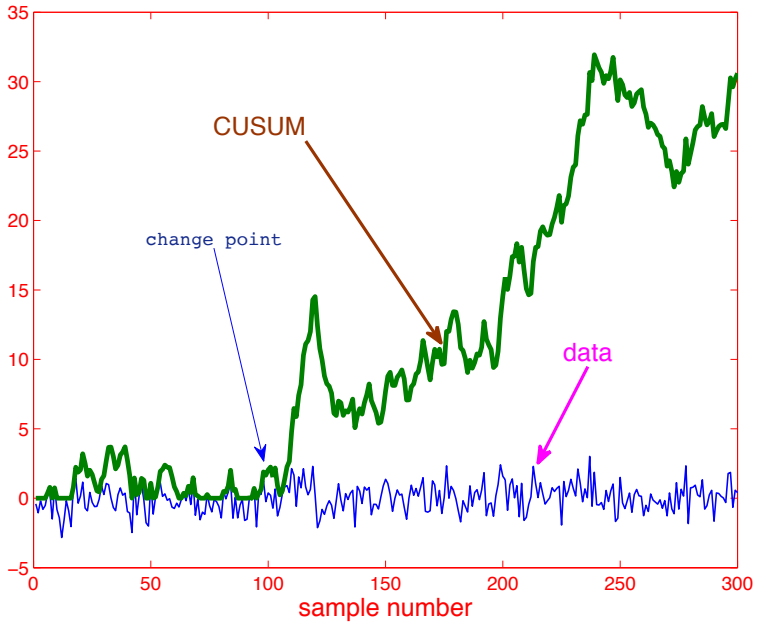

Figure 5: A typical example of the Page test finding a change from $\mathcal{N}(-0.2, 1)$ to $\mathcal{N}(+0.2, 1)$ at time $n = 100$.

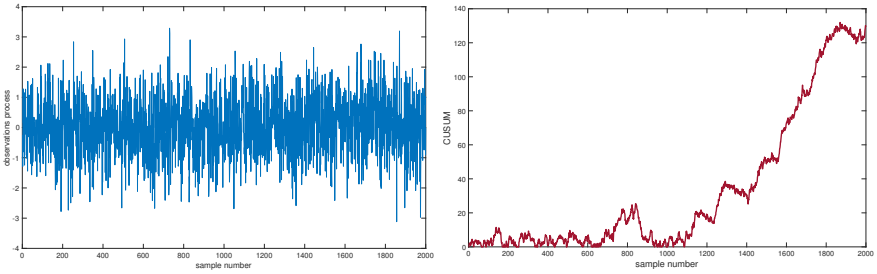

Figure 6: Similar to figure 5, except a change from $\mathcal{N}(-0.1, 1)$ to $\mathcal{N}(+0.1, 1)$ at time $n = 1000$. smaller changes take longer.

$$N_b(\theta) \quad = \quad \mathcal{E}\{\text{number of samples in Page test}\} \tag{95}$$

where the condition on θ is obvious. Now we have

$$\begin{aligned} N_b(\theta) \quad = \quad & \sum_{m=0}^{\infty} \mathcal{E}\{\text{number of samples given m resets at } b\} \\ & \times Pr(m \text{ resets at } b \text{ before ending at } h) \end{aligned} \tag{96}$$

$$= \sum_{m=0}^{\infty} \left[M_h(\theta) + mM_b(\theta)\right] P(\theta)^m (1 - P(\theta)) \tag{97}$$

$$= M_h(\theta) + M_b(\theta) \frac{P(\theta)}{(1 - P(\theta))} \tag{98}$$

$$= \frac{M(\theta)}{1 - P(\theta)} \tag{99}$$

since by definition

$$M(\theta) \;=\; P(\theta)M_b(\theta) \;+\; (1 - P(\theta))M_h(\theta) \tag{100}$$

We are interested in $N_b(\theta)|_{b=0}$ so that this repeated Wald test is the same as the Page test.

From Wald's Identity, given there is a negative bias on the test under $\mathcal{H}$, we have

$$M(\theta) \;=\; \frac{P(\theta)b + (1 - P(\theta))h}{\mathcal{E}\{g(x_i)|\theta\}} \tag{101}$$

Unfortunately this means $N_b(\theta)|_{b=0}$ is not well-defined, since

$$\lim_{b \to 0} P(\theta) = 1 \tag{102}$$

means we have

$$\lim_{b \to 0} M(\theta) = 0 \tag{103}$$

and (99) fails.

Now we can write

$$N_b(\theta) \;=\; \frac{P(\theta)b}{(1 - P(\theta))\mathcal{E}\{g(x_i)|\theta\}} \;+\; \frac{h}{\mathcal{E}\{g(x_i)|\theta\}} \tag{104}$$

from insertion of (101) into (99). We examine Wald's Identity

$$\mathcal{E}\{e^{ts_M} \left[\phi_\theta(t)\right]^{-M} |\theta\} \;=\; 1 \tag{105}$$

where, to repeat

$$s_M \;=\; \text{the value of the SPRT at termination} \tag{106}$$

$$M \;=\; \text{the number of samples to the SPRT's termination} \tag{107}$$

$$\phi_\theta(t) \;=\; \mathcal{E}\{e^{tg(x_i)}|\theta\} \tag{108}$$

in which we have adjusted notation from (53) to reflect what we need now.

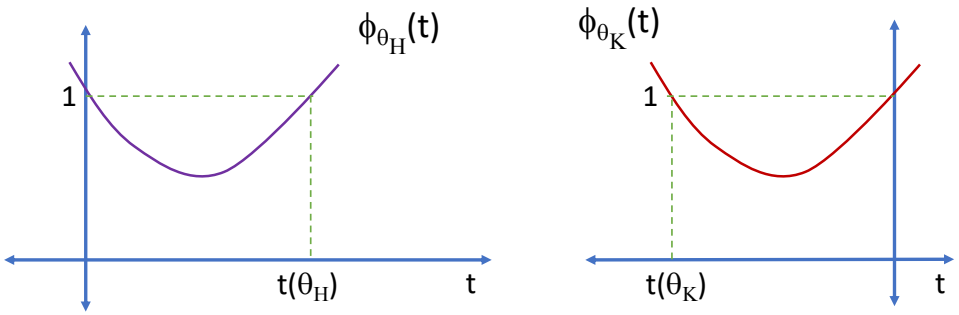

Figure 7: The moment generating functions from (108) for θ_H and θ_K. What is of interest are the "roots"' (where $\phi_\theta(t) = 1$) under the two hypotheses.

Consider this moment generating function $\phi_\theta(t)$. We have

$$\begin{aligned}
\phi_\theta(0) &= \mathcal{E}\{e^{tg(x_i)}|\theta\}\Big|_{t=0} = 1 && (109)\\
\dot{\phi}_\theta(0) &= \mathcal{E}\{g(x_i)e^{tg(x_i)}|\theta\}\Big|_{t=0} = \mathcal{E}\{g(x_i)|\theta\} && (110)\\
\ddot{\phi}_\theta(0) &= \mathcal{E}\{g(x_i)^2e^{tg(x_i)}|\theta\}\Big|_{t=0} \geq 0 && (111)\\
\lim_{t\to\infty} \phi_{\theta_H}(t) &= \lim_{t\to\infty} \mathcal{E}\{e^{tg(x_i)}|\theta_H\} = \infty && (112)\\
\lim_{t\to-\infty} \phi_{\theta_K}(t) &= \lim_{t\to-\infty} \mathcal{E}\{e^{tg(x_i)}|\theta_K\} = \infty && (113)
\end{aligned}$$

To justify (112), first admit that (112) must be true if $Pr(g(x_i) > 0|\theta_H) > 0$; this is easy to see. But if $Pr(g(x_i) > 0|\theta_H) = 0$ then either $Pr(g(x_i) > 0|\theta_K) = 0$ as well (and this makes little sense given the CUSUM structure) or else the Page procedure needs to modified to accept it. The modification would be that – since $g(x_i) > 0$ can only happen under $\mathcal{K}$ – then the testing strategy amounts to waiting for an x_i such that $g(x_i) > 0$ and then declaring a change. There is nothing intrinsically wrong with this strategy, but it eviscerates the analysis[4] we are about to perform, in which we attempt to find a relation between $\bar{T}$ and the threshold h. Similar comments apply to (113): we have to assume that $Pr(g(x_i) < 0|\theta_K) > 0$ else use some different sort of analysis.

This yields the moment generating function behaviors seen in figure 7 and illustrates that there must exist at least one non-zero "root" (where $\phi_\theta(t) = 1$), a positive root under $\mathcal{H}$ since $\mathcal{E}\{g(x_i)|\theta_H\} < 0$ and a negative

[4]That is, we have to assume that $Pr(g(x_i) > 0|\theta_H) > 0$, with the knowledge that otherwise $\bar{T} = \infty$ and $\bar{D}$ is given by the mean of a geometric pmf.

root under $\mathcal{K}$ since $\mathcal{E}\{g(x_i)|\theta_K\} > 0$. Let us call such a root $t(\theta)$. Since the SPRT can only end in b or h, we evaluate (105) at $t(\theta)$ and get

$$e^{t(\theta)b}P(\theta) \; + \; e^{t(\theta)h}(1 - P(\theta)) \; = \; 1 \tag{114}$$

We can solve this for $P(\theta)$ as

$$P(\theta) \;\; = \;\; \frac{1 - e^{t(\theta)h}}{e^{t(\theta)b} - e^{t(\theta)h}} \tag{115}$$

$$\frac{P(\theta)}{1 - P(\theta)} \;\; = \;\; \frac{1 - e^{t(\theta)h}}{e^{t(\theta)b} - 1} \tag{116}$$

We can insert this to (104) and get

$$N_b(\theta) \; = \; \frac{1}{\mathcal{E}\{g(x_i)|\theta\}} \left[(1 - e^{t(\theta)h}) \frac{b}{e^{t(\theta)b} - 1} \; + \; h \right] \tag{117}$$

Via L'Hopital we know

$$\lim_{b \to 0} \left\{ \frac{b}{e^{t(\theta)b} - 1} \right\} \; = \; \frac{1}{t(\theta)} \tag{118}$$

hence

$$N_0(\theta) \;\; = \;\; \lim_{b \to 0} \{N_b(\theta)\} \tag{119}$$

$$\;\; = \;\; \frac{1 + ht(\theta) - e^{ht(\theta)}}{t(\theta)\mathcal{E}\{g(x_i)|\theta\}} \tag{120}$$

which is the fundamental equation describing Page performance.

Now let us assume that h is large – that would be what we want, to achieve a large $\bar{T}$. Then recalling that $\mathcal{E}\{g(x_i)|\theta_H\} < 0$ and hence $t(\theta_H) > 0$, we approximate (120) under $\mathcal{H}$ as

$$\bar{T} \; \approx \; \frac{-e^{ht(\theta_H)}}{t(\theta_H)\mathcal{E}\{g(x_i)|\theta_H\}} \tag{121}$$

in which the negativity of $\mathcal{E}\{g(x_i)|\theta_H\}$ should be recalled. Further, since $\mathcal{E}\{g(x_i)|\theta_K\} > 0$ and hence $t(\theta_K) < 0$, we approximate (120) under $\mathcal{K}$ as

$$\bar{D} \; \approx \; \frac{h}{\mathcal{E}\{g(x_i)|\theta_K\}} \tag{122}$$

If we therefore define

$$\eta \equiv \frac{\log(\bar{T})}{\bar{D}} \tag{123}$$

$$\approx \frac{ht(\theta_H) - \log(t(\theta_H)) - \log(-\mathcal{E}\{g(x_i)|\theta_H\})}{\left(\frac{h}{\mathcal{E}\{g(x_i)|\theta_K\}}\right)} \tag{124}$$

$$\approx t(\theta_H)\mathcal{E}\{g(x_i)|\theta_K\} \tag{125}$$

This is truly remarkable: the average time between false alarms is (asymptotically, as $h \to \infty$) exponential in the average delay to detection. That is, if we have a test with $\bar{T} = 10^3$ and $\bar{D} = 10$, we could get $\bar{T} = 10^6$ and $\bar{D} = 20$.

In the case that

$$g(x) = \log\left(\frac{f(x|\theta_K)}{f(x|\theta_H)}\right) \tag{126}$$

we have the equation defining $t(\theta_H)$ as

$$1 = \mathcal{E}\{e^{t(\theta_H)g(x)}\} \tag{127}$$

$$= \int \left(\frac{f(x|\theta_K)}{f(x|\theta_H)}\right)^{t(\theta_H)} f(x|\theta_H)dx \tag{128}$$

$$= \int f(x|\theta_K)^{t(\theta_H)} f(x|\theta_H)^{1-t(\theta_H)} dx \tag{129}$$

$$\tag{130}$$

which means $t(\theta_H) = 1$ and

$$\eta = \mathcal{E}\{g(x_i)|\theta_K\} \tag{131}$$

$$= \int \log\left(\frac{f(x|\theta_K)}{f(x|\theta_H)}\right) f(x|\theta_K)dx \tag{132}$$

which is the "divergence" – one term in the J-divergence.

It should be noted that Wald's approximations for the run-lengths, on which the Page analysis is based, are predicated on the assumption that the "excess over the boundary" is negligible relative to the boundary. In the Page case, while this may be true for a crossing of the upper boundary h, it is manifestly not so for the lower boundary $b \to 0^-$. As a result, there is some inaccuracy in the analysis. More accurate expressions have been derived by Siegmund ("Siegmund's Correction Terms") – while these are based on some rather gross approximations, the results seem to be fairly accurate.

3.3 An Example

Not too surprisingly, let us look at

$$f(x|\theta) \;=\; \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} \tag{133}$$

where $\theta_H = 0$ and $\theta_K = \mu$. To explore the role of the "bias," let us write

$$g(x) \;=\; x - \gamma \tag{134}$$

Now

$$\begin{aligned}
\phi_{\theta_H}(t) &= \mathcal{E}\{e^{tg(x)}|\theta_H\} && (135)\\
&= \int e^{t(x-\gamma)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx && (136)\\
&= \int \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} dx e^{-t\gamma + t^2/2} && (137)\\
&= e^{-t\gamma + t^2/2} && (138)
\end{aligned}$$

Hence if we seek

$$\phi_{\theta_H}(t(\theta_H)) \;=\; 1 \tag{139}$$

we have

$$\begin{aligned}
-t(\theta_H)\gamma + t(\theta_H)^2/2 &= 0 && (140)\\
t(\theta_H) &= 2\gamma && (141)
\end{aligned}$$

as the non-zero root. Continuing, we have

$$\begin{aligned}
\eta &= t(\theta_H)\mathcal{E}\{g(x_i)|\theta_K\} && (142)\\
&= 2\gamma(\mu - \gamma) && (143)
\end{aligned}$$

and naturally this is optimized for

$$\gamma \;=\; \mu/2 \tag{144}$$

meaning that $g(x)$ is the log-likelihood ratio, and if we used that we would get $\eta = \mu^2/2$.

Recall that from (121) and (122) we have

$$\bar{T} \;\approx\; \frac{-e^{ht(\theta_H)}}{t(\theta_H)\mathcal{E}\{g(x_i)|\theta_H\}} \tag{145}$$

and

$$\bar{D} \;\approx\; \frac{h}{\mathcal{E}\{g(x_i)|\theta_K\}} \tag{146}$$

Suppose $\mu = 1$ and it is desired to have $\bar{T} = 10^6$. If we use the optimal *bias* $\gamma = \mu/2$ then we have

$$10^6 \;=\; \frac{-e^h}{-1/2} \tag{147}$$

or $h = 13.1$. The corresponding

$$\bar{D} \;\approx\; \frac{h}{\mathcal{E}\{g(x_i)|\theta_K\}} \;=\; \frac{13.1}{1/3} \;=\; 27 \tag{148}$$

Not so bad.

But suppose we set $\gamma = \mu/4$, a too-small bias that presumably comes from underestimating the size of the change. Then we have $t(\theta_H) = 1/2$ and hence (121) gives us

$$10^6 \;=\; \frac{-e^{h/2}}{(1/2)(-1/4)} \tag{149}$$

or $h = 23.5$, and thence (122) yields

$$\bar{D} \;\approx\; \frac{h}{\mathcal{E}\{g(x_i)|\theta_K\}} \;=\; \frac{23.5}{0.75} \;=\; 31 \tag{150}$$

3.4 The Local Case

Without loss of generality consider we have

$$\begin{aligned} \mathcal{H}: &\quad x_i \;=\; \nu_i \;-\; \theta \\ \mathcal{K}: &\quad x_i \;=\; \nu_i \;+\; \theta \end{aligned} \tag{151}$$

where ν_i is *iid* with pdf $f(\cdot)$ and in which the update $g(x_i)$ such that

$$\int g(x) f(x) dx \;=\; 0 \tag{152}$$

is used in the Page procedure. Recall that we have

$$\eta(\theta) \;=\; t(\theta)\mathcal{E}\{g(x_i)|\theta\} \tag{153}$$

in which

$$\mathcal{E}\{e^{t(\theta)g(x)}\} \;=\; 1 \tag{154}$$

Let us use the Taylor strategy:

$$\eta(\theta) \approx \eta(0) + \theta\dot{\eta}(0) + \frac{1}{2}\theta^2\ddot{\eta}(0) \tag{155}$$

Now

$$\eta(0) = t(\theta)\mathcal{E}\{g(x_i)|\theta\}|_{\theta=0} = 0 \tag{156}$$

by assumption about $g(x)$; and it can also be shown that $t(0) = 0$; so $\eta(0)$ is especially small.

Continuing, we have (let us take $\mathcal{H}$ for concreteness)

$$\dot{\eta}(0) = \dot{t}(\theta)\int f(x-\theta)g(x)dx - t(\theta)\int \dot{f}(x-\theta)g(x)dx\Big|_{\theta=0} \tag{157}$$

$$= (\dot{t}(0)\times 0) + \left(0\times\int \dot{f}(x)g(x)dx\right) \tag{158}$$

$$= 0 \tag{159}$$

Hence what remains is

$$\ddot{\eta}(0) = \ddot{t}(0)\int f(x)g(x)dx - \dot{t}(0)\int \dot{f}(x)g(x)dx t(\theta) - \dot{t}(0)\int \dot{f}(x)g(x)dx + t(0)\int \ddot{f}(x)g(x)dx \tag{160}$$

$$= -2\dot{t}(0)\int \dot{f}(x)g(x)dx \tag{161}$$

and clearly we need $\dot{t}(\theta)$.

We write

$$1 = \int e^{t(\theta)g(x)}f(x+\theta)dx \tag{162}$$

and differentiate to get

$$0 = \dot{t}(\theta)\int e^{t(\theta)g(x)}g(x)f(x+\theta)dx + \int e^{t(\theta)g(x)}\dot{f}(x+\theta)dx \tag{163}$$

$$= \dot{t}(0)\int g(x)f(x)dx + \int \dot{f}(x)dx \tag{164}$$

which is actually telling us that $0 = 0$ – true but not especially useful. So let us bravely differentiate again:

$$0 = \ddot{t}(0)\int e^{t(0)g(x)}g(x)f(x)dx + \dot{t}(0)^2\int e^{t(0)g(x)}g(x)^2f(x)dx + \dot{t}(0)\int e^{t(0)g(x)}g(x)\dot{f}(x)dx + \dot{t}(0)\int e^{t(0)g(x)}g(x)\dot{f}(x)dx + \int e^{t(0)g(x)}\ddot{f}(x)dx \tag{165}$$

$$= \dot{t}(0)^2\int g(x)^2f(x)dx + 2\dot{t}(0)\int g(x)\dot{f}(x)dx \tag{166}$$

or

$$\dot{t}(0) \;=\; \frac{-2\int g(x)\dot{f}(x)dx}{\int g(x)^2 f(x)dx} \tag{167}$$

Now we get

$$\begin{aligned}
\eta &\approx \frac{1}{2}\theta^2\ddot{\eta}(0) && (168)\\
&= \frac{1}{2}\theta^2(-2\dot{t}(0)\int \dot{f}(x)g(x)dx) && (169)\\
&= \frac{1}{2}\theta^2(-2\frac{-2\int g(x)\dot{f}(x)dx}{\int g(x)^2 f(x)dx}\int \dot{f}(x)g(x)dx && (170)\\
&= 2\theta^2\frac{\left(\int g(x)\dot{f}(x)dx\right)^2}{\int g(x)^2 f(x)dx} && (171)\\
&= 2E(g,f)^2 && (172)
\end{aligned}$$

which basically closes the circle: once again, we see efficacy.

3.5 The Shiryaev Approach

This amounts to the Bayesian version of the Page procedure. That is, we assume an underlying binary random process $\{\theta_i\}$ such that $\theta_i \in \{\theta_H, \theta_K\}$ and

$$Pr(\theta_i|\theta_{i-1}) \;=\; \begin{pmatrix} 1-\rho & \rho \\ 0 & 1 \end{pmatrix} \tag{173}$$

meaning that the hypothesis process is a non-recurrent Markov chain: once the change happens, it never switches back. Clearly the change process is geometrically distributed, with expected number of samples before a change $1/\rho$.

It is easy to derive a sufficient statistic

$$\begin{aligned}
P_i &= Pr(\theta_i = \theta_K|\{x_j\}_{j=1}^{i}) && (174)\\
&= c\Big[Pr(\theta_i = \theta_K|\theta_{i-1} = \theta_K)Pr(\theta_{i-1} = \theta_K|\{x_j\}_{j=1}^{i-1})f(x_i|\theta_K) && (175)\\
&\qquad +Pr(\theta_i = \theta_K|\theta_{i-1} = \theta_H)Pr(\theta_{i-1} = \theta_H|\{x_j\}_{j=1}^{i-1})f(x_i|\theta_K)\Big]\\
&= c\left[P_{i-1}f(x_i|\theta_K) + \rho(1-P_{i-1})f(x_i|\theta_K)\right] && (176)\\
&= c\left[P_{i-1} + \rho(1-P_{i-1})\right]f(x_i|\theta_K) && (177)
\end{aligned}$$

and examination of the sequence $\{P_i\}$ with appropriate threshold to declare a "detection" makes sense.

Some people prefer a likelihood ratio formulation. It is easy to show directly, as with (177), that we can write

$$\begin{aligned} 1 - P_i &= Pr(\theta_i = \theta_H | \{x_j\}_{j=1}^{i}) && (178) \\ &= c\Big[Pr(\theta_i = \theta_H | \theta_{i-1} = \theta_K) Pr(\theta_{i-1} = \theta_K | \{x_j\}_{j=1}^{i-1}) f(x_i|\theta_H) && (179) \\ &\qquad + Pr(\theta_i = \theta_H | \theta_{i-1} = \theta_H) Pr(\theta_{i-1} = \theta_H | \{x_j\}_{j=1}^{i-1}) f(x_i|\theta_H)\Big] \\ &= c(1 - P_{i-1})(1 - \rho) f(x_i|\theta_H) && (180) \end{aligned}$$

This term c is a normalizing constant, and its value doesn't much matter unless (177) is being used directly. It can be calculated as

$$\begin{aligned} 1/c &\equiv f(\{x_j\}_{j=1}^{i}) && (181) \\ &= [P_{i-1} + \rho(1 - P_{i-1})]\, f(x_i|\theta_K) \\ &\qquad + (1 - P_{i-1})(1 - \rho) f(x_i|\theta_K) && (182) \end{aligned}$$

So with

$$T_i(x) = \frac{P_k}{1 - P_i} \tag{183}$$

and

$$L(x_i) = \frac{f(x_i|\theta_K)}{f(x_i|\theta_H)} \tag{184}$$

we have

$$\begin{aligned} T_i(x) &= \frac{c\,[P_{i-1} + \rho(1 - P_{i-1})]\, f(x_i|\theta_K)}{c(1 - P_{i-1})(1 - \rho) f(x_i|\theta_H)} && (185) \\ &= \left(\frac{T_{i-1}(x) + \rho}{1 - \rho}\right) L(x_i) && (186) \end{aligned}$$

A logarithm form is often convenient:

$$\begin{aligned} t_i(x) &\equiv \log(T_i(x)) && (187) \\ &= \log(e^{t_{i-1}(x)} + \rho) - \log(1 - \rho) + \log(L(x_i)) && (188) \end{aligned}$$

The Shiryaev-Roberts form of the test statistic assumes ρ is vanishingly small and defines

$$\begin{aligned} R_i(x) &= \frac{T_i(x)}{\rho} && (189) \\ &= (R_{i-1}(x) + 1) L(x_i) && (190) \end{aligned}$$

Both (188) and (190) are CUSUM-like.

Two important notes. First, initialization of (177)/(182), (188) or (190) is a practical concern, but generally not very important unless an extreme value is selected. Second, while all this analysis has assumed that the observations process $\{x_i\}$ is conditionally independent, there is, unlike for the CUSUM's optimality, no real need for that. Of course (177)/(182), (188) and (190) all need to be modified if independence is not the case; but that is not difficult. The Shiryaev approach is especially valuable in non-independent situations; for example, consider that $\{\theta_i\}$ is the target-existence variable for possible track deletion, but the observations process $\{x_i\}$ is the detection process of the target, and $\{P_d(i)\}$ is itself a Markov process if the target undergoes periods of low-detectability.

ECE 6124
Signal Detection
Fundamentals of Radar and Sonar

Peter Willett

Fall 2018

1 The Signal

The usual time-domain model for radar (and to a more relaxed extent, sonar) is that a signal

$$s(t) \;=\; \Re\left\{A_c p(t) e^{j2\pi f_c t}\right\} \tag{1}$$

is transmitted, in which $p(t)$ is some canonical pulse shape and the phase and time references are for convenience assumed to be zero. For example, we could have

$$p(t) \;=\; u(t) - u(t-T) \tag{2}$$

as the *baseband pulse*, in which case the transmitted signal is just

$$s(t) \;=\; \begin{cases} A_c \cos(2\pi f_c t) & 0 \le t < T \\ 0 & \text{else} \end{cases} \tag{3}$$

Alternatively we could have

$$p(t) \;=\; [u(t) - u(t-T)] e^{j2\pi(-f_\Delta t + f_\Delta t^2/T)/2} \tag{4}$$

in which case

$$s(t) \;=\; \begin{cases} A_c \cos(2\pi(f_c t + (-f_\Delta t + f_\Delta t^2/T)/2)) & 0 \le t < T \\ 0 & \text{else} \end{cases} \tag{5}$$

in which case the transmitted pulse is an up-chirp (or LFM). To see this, note that the instantaneous frequency of a sinusoidal signal

$$y(t) \;=\; A\cos(\phi(t)) \tag{6}$$

is agreed to be

$$f(t) \;=\; \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{7}$$

which in the case of (5) is

$$f(t) \;=\; f_c - f_\Delta/2 + f_\Delta t/T \tag{8}$$

over the range $0 \leq t < T$: that is, a linear function from $-f_\Delta/2$ to $f_\Delta/2$. Other signals of interest include coded pulses, "noise radar" pulses and multiple pulses of all such types.

Let us briefly consider the effect of the relative position of transmitter and target.

Delay. If the target is located at a distance d from the transmitter, the received signal (absent noise) is $s(t - 2d/c)$, in which c is the speed of propagation in the medium. Examples include a rough order of magnitude $c \approx 1500m/s$ for sound in water (sonar), $c \approx 300m/s$ for sound in air (acoustics) and $c \approx 3 \times 10^8 m/s$ for electromagnetic radiation (radar). The factor of 2 is due to the round trip, assuming colocated[1] transmitter and receiver.

Power and Energy. In a situation of *spherical spreading* (as generally the case with radar) the assumption is that all spherical shells surrounding the transmitter, contain the same power. This would mean that the "area" by which a given target collects transmitted energy owns a proportion of that shell that varies as $1/d^2$. Hence, since the target, when it re-radiates this power back to the (colocated) receiver is subject to the same law. The result is that power is proportional to $1/d^4$ in radar. Whether this applies to sonar or not depends on target range and water column depth (and many other things too); but it is often assumed that the range is much larger than depth and hence power spreading is "planar." As such, it is common to assume an aggregate $1/d^2$ law in sonar – seemingly more favorable than radar. Energy is of course power times time; so the energy collected is proportional to pulse-length. That is, for radar (at least) $SNR \propto T/d^4$; and for sonar $SNR \propto T/d^2$ (usually).

Coded pulses were mentioned earlier, and of special interest are the Barker coded pulses for which

$$p(t) \;=\; \sum_{n=0}^{N-1} \beta_n q(t - k\tau) \tag{9}$$

and in which $q(t)$ is nominally a square pulse of "chip" length τ but can be otherwise. The sequence $\{\beta_n\} \in \{-1, 1\}$ is ideally chosen such that

$$\frac{1}{N}\left|\sum_n \beta_n \beta_{n+m}\right| \;\leq\; \frac{1}{N} \tag{10}$$

[1]Bistatic operation is actually often quite favorable, especially for reasons of covertness in which the transmitted may be an expendable "dumb" device but the receiver a high-value asset.

unless $m = 0$ (in which case the sum is unity). This means that such (Barker) pulses have autocorrelation sidelobes as small as possible, given their structure. There are only four such sequences known. The cases $N = 2$ & $N = 3$ (respectively $\{+1, -1\}$ and $\{+1, +1, -1\}$) are largely trivial. The interesting cases[2] are $N = 7$

$$\beta_n \;=\; \{+1, +1, +1, -1, -1, +1, -1\} \tag{11}$$

and $N = 11$

$$\beta_n \;=\; \{+1, +1, +1, -1, -1, -1, +1, -1, -1, +1, -1\} \tag{12}$$

The Barker codes can be compounded: that is, the pulses $q(t)$ can themselves be Barker-coded with correspondingly shorter chip lengths.

The fundamental pulse $p(t)$ can also be *windowed* (or "shaded" or "tapered") to reduce sidelobes. Whereas for digital signal processing filter designs the window function in the time domain is measured in the frequency-domain – sharp transition band and low sidelobes – here good performance is evaluated in terms of the ambiguity function, which is discussed shortly but is essentially the signal convolved with its time-reversed version. A good ambiguity function would have low sidelobes and sharp transition band; but it would also have a narrow "passband" (for resolution) and little variation around the peak up to the transition (little SNR loss). Many systems use Taylow weighting for this.

2 Detection

2.1 The Nonfluctuating Case

In the absence of distortion – and hence this applies more to radar than to sonar, where distortion is endemic – the observed signal[3] is

$$\mathcal{H} : x(t) \;=\; \nu(t) \tag{13}$$

$$\mathcal{K} : x(t) \;=\; Ap(t)\cos(2\pi f_c(t - t_0 - t^{\Delta})) \;+\; \nu(t) \tag{14}$$

in which $\nu(t)$ is additive white Gaussian noise with power spectral density $N_0/2$. In (14), t_0 is the reference time – the time corresponding to the

[2] To see why things are interesting look at the matched-filter response for $\{+1, -1, +1, -1\}$, which is seemingly good.

[3] This is sometimes called the nonfluctuating case: fluctuations can be cause by propagation disturbances, but more likely come from a target that contains multiple reflectors: since their geometry can change rapidly from constructive to destructive interference, the target is said to fluctuate.

location of the target that is being tested for – and t^{Δ} is small offset from that. To be concrete, suppose we have $f_c = 3GHz$, $T = 1\mu s$ and we test for targets every μs, which corresponds to testing the hypothesis that the target is at (say) 1000m, 1150m, 1300m, etc. But suppose the true target is at a distance 1156m: this would indicate that the target is 6m mismatched to the test at 1150m, and hence $t^{\Delta} = 2(6m)/(3 \times 10^8 m/s) = 0.04\mu s$. Now, for notational ease we set $t_0 = 0$, and write

$$\begin{aligned} \mathcal{H} : x(t) &= \nu(t) && (15) \\ \mathcal{K} : x(t) &= Ap(t)\cos(2\pi f_c t - 2\pi f_c t^{\Delta}) \ + \ \nu(t) && (16) \end{aligned}$$

We see that (in this nominal example) $f_c t^{\Delta} = (3 \times 10^9 Hz)(4 \times 10^{-8} s) = 120$ – that is, even this small 6m offset from the 1150m test point corresponds to 120 periods of the carrier waveform. It is for this reason that we write

$$\begin{aligned} \mathcal{H} : x(t) &= \nu(t) && (17) \\ \mathcal{K} : x(t) &= Ap(t)\cos(2\pi f_c t + \psi) \ + \ \nu(t) && (18) \end{aligned}$$

in which ψ is a phase assumed uniformly distributed on $(0, 2\pi)$, since as indicated with the previous notional example, the wavelength of a radar waveform is $\lambda = c/f_c$. If $f_c = 3GHz$ we have $\lambda = 10cm$, and there is no reasonable way to match a target's location with this accuracy.

Without loss of generality[4] we consider the "baseband" version of this observations process:

$$\begin{aligned} x^c(t) &= \mathcal{LPF}\left\{x(t)2\cos(2\pi f_c t)\right\} && (19) \\ x^s(t) &= \mathcal{LPF}\left\{x(t)2\sin(2\pi f_c t)\right\} && (20) \end{aligned}$$

in which it is assumed that $p(t)$ has a bandwidth B that corresponds to the cutoff frequency of the low-pass filter. We could reconstruct

$$x(t) \ = \ x^c(t)\cos(2\pi f_c t) \ - \ x^s(t)\sin(2\pi f_c t) \qquad (21)$$

and we note that these $x_c(t)$ and $x_s(t)$ are know as the *in-phase* and *quadrature* ("I & Q") channels – and radar systems really do use these[5] in their hardware. We have under $\mathcal{K}$ that we can write

$$\begin{aligned} x^c(t) &= Ap(t)\cos(\psi) \ + \ \nu^c(t) && (22) \\ x^s(t) &= Ap(t)\sin(\psi) \ + \ \nu^s(t) && (23) \end{aligned}$$

[4]The statistical modeling theory is discussed in Communication Theory classes. It is standard but too lengthy to repeat in an inline discussion. However, the relevant notes are provided separately.

[5]That is, if you looked at the electronics these signals would really actually exist and pass the "oscilloscope" test: you could touch a probe to a wire and see them on the scope.

where $\nu_c(t)$ and $\nu_s(t)$ are independent with flat power-spectral densities N_0 on $(-B, B)$, and zero elsewhere.

We could (and probably should) use the Karhunen-Loève formulation for this, but it is clearer to assume that the signals are ideally lowpass-filtered sampled at the Nyquist frequency such that the (filtered white) noise samples are independent. That is, we have

$$x_i^c = Ap_i \cos(\psi) + \nu_i^c \tag{24}$$
$$x_i^s = Ap_i \sin(\psi) + \nu_i^s \tag{25}$$

in which ψ remains uniform on $(0, 2\pi)$ and the ν's are all *iid* Gaussian with means zero and variances $\sigma^2 = N_0 f_s$ in which f_s is the (Nyquist) sampling rate. We have

$$\begin{aligned} & f(x_c, x_s|\mathcal{K}) \\ &= \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i^c - Ap_i\cos(\psi))^2 + (x_i^s - Ap_i\sin(\psi))^2]}d\psi \qquad (26) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i^c)^2 + (x_i^s)^2 + A^2p_i^2]} \\ &\quad \times \frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{\sigma^2}\left(\sum_{i=1}^n[x_i^c Ap_i]\cos(\psi) + \sum_{i=1}^n[x_i^s Ap_i]\sin(\psi)\right)}d\psi \qquad (27) \end{aligned}$$

Via the standard change of variables we define

$$r\cos(\theta) \equiv \sum_{i=1}^n [x_i^c p_i] \quad (\equiv u) \tag{28}$$
$$r\sin(\theta) \equiv \sum_{i=1}^n [x_i^s p_i] \quad (\equiv v) \tag{29}$$

hence (27) becomes

$$\begin{aligned} f(x_c, x_s|\mathcal{K}) &= \left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i^c)^2 + (x_i^s)^2 + A^2p_i^2]} \\ &\quad \times \frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{\sigma^2}(Ar\cos(\theta)\cos(\psi) + Ar\sin(\theta)\sin(\psi))}d\psi \qquad (30) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i^c)^2 + (x_i^s)^2 + A^2p_i^2]} \\ &\quad \times \frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{\sigma^2}Ar\cos(\psi-\theta)}d\psi \qquad (31) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n[(x_i^c)^2 + (x_i^s)^2 + A^2p_i^2]} I_0(Ar) \qquad (32) \end{aligned}$$

in which $I_0(\cdot)$ denotes a Bessel function. Since we also have

$$f(x_c, x_s|\mathcal{H}) \;=\; \left(\frac{1}{2\pi\sigma^2}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(x_i^c)^2+(x_i^s)^2]} \tag{33}$$

we have, by forming the likelihood ratio, dropping constants and realizing the monotonicity of $I_0(\cdot)$, that a sufficient statistic for testing is

$$\begin{aligned}
\lambda(x_c, x_s) \;&=\; r^2 \\
&\equiv\; u^2 + v^2 && (34)\\
&=\; \left(\sum_{i=1}^n x_i^c p_i\right)^2 + \left(\sum_{i=1}^n x_i^s p_i\right)^2 && (35)\\
&\propto\; \left|\int_0^T \tilde{x}(t)p(t)^* dt\right|^2 && (36)\\
&=\; \left(\int_0^T x(t)p(t)\cos(2\pi f_c t)dt\right)^2 \;+\; \left(\int_0^T x(t)p(t)\sin(2\pi f_c t)dt\right)^2 && (37)
\end{aligned}$$

in which (36) uses the complex baseband representation of the signal

$$\tilde{x}(t) \;=\; x^c(t) + jx^s(t) \tag{38}$$

as discussed in the separate notes on narrowband signals and noise.

Under $\mathcal{H}$, u and v are independent and Gaussian, with means zero and total variances

$$\sigma^2 \;=\; \frac{N_0}{2}\int_0^T p(t)^2 dt \tag{39}$$

We thus have

$$\begin{aligned}
\alpha \;&=\; Pr(T > \tau\sigma^2|\mathcal{H}) && (40)\\
&=\; \int_{u^2+v^2>\tau\sigma^2} \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[u^2+v^2]} du dv && (41)\\
&=\; \int_{\sqrt{\tau}\sigma}^{\infty}\int_0^{2\pi} \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr d\theta && (42)\\
&=\; e^{-\tau/2} && (43)
\end{aligned}$$

meaning that the test statistic has an exponential density under $\mathcal{H}$. In (42) we used (28), (29) and took the Jacobian.

Under $\mathcal{K}$, u and v are independent and Gaussian, with total variance σ^2 as in (39) and means

$$\mathcal{E}\{u|\mathcal{K}\} \;=\; \int_0^T Ap(t)\cos(2\pi f_c t + \psi)p(t)\cos(2\pi f_c t)dt \tag{44}$$

$$
\begin{aligned}
&= \frac{A}{2}\int_0^T p(t)^2 dt \cos(\psi) && (45)\\
&\equiv \rho\cos(\psi) && (46)\\
\mathcal{E}\{v|\mathcal{K}\} &= \int_0^T Ap(t)\cos(2\pi f_c t+\psi)p(t)\sin(2\pi f_c t)dt && (47)\\
&= \frac{A}{2}\int_0^T p(t)^2 dt \sin(\psi) && (48)\\
&\equiv \rho\sin(\psi) && (49)
\end{aligned}
$$

in which

$$\rho \equiv \frac{A}{2}\int_0^T p(t)^2 dt \quad (50)$$

We get

$$
\begin{aligned}
\beta &= Pr(T > \tau\sigma^2|\mathcal{K}) && (51)\\
&= \int_{T>\tau\sigma^2} \frac{1}{\pi\sigma^2} e^{-\frac{1}{\sigma^2}[(u-\rho\cos(\psi))^2+(v-\rho\sin(\psi))^2]} du dv && (52)\\
&= \int_{\sqrt{\tau}\sigma}^{\infty}\int_0^{\pi} \frac{r}{\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(r\cos(\theta)-\rho\cos(\psi))^2+(r\sin(\theta)-\rho\sin(\psi))^2]} dr d\theta && (53)\\
&= \int_{\sqrt{\tau}\sigma}^{\infty} \frac{2r}{\sigma^2} e^{-\frac{r^2+\rho^2}{\sigma^2}} I_0(r\rho/\sigma^2) dr && (54)\\
&= \int_{\sqrt{\tau}}^{\infty} Re^{-\frac{R^2+(\rho/\sigma)^2}{2}} I_0(R\rho/\sigma) dR && (55)\\
&\equiv \mathcal{Q}\left(\frac{\rho}{\sigma}, \sqrt{\tau}\right) && (56)
\end{aligned}
$$

To get to (55) we used $R = r/\sigma$, and the integrand in (55) is called the *Ricean* pdf. The function in (56) is called *Marcum's Q-function* and it is tabulated if not especially pretty. The SNR term in (56) is

$$
\begin{aligned}
SNR &\equiv (\rho/\sigma)^2 && (57)\\
&= \frac{A^2/4\left(\int_0^T p(t)^2 dt\right)^2}{N_0/2\int_0^T p(t)^2 dt} && (58)\\
&= \frac{A^2}{2N_0}\int_0^T p(t)^2 dt && (59)\\
&= \frac{E}{N_0} && (60)
\end{aligned}
$$

in which E is the radar return energy, the integral of the square of (3).

2.2 The Swerling I Model

Equation (16) represents a "point" target: a single reflection. This may be appropriate either for a waveform with very low resolution or a target that is a small metal ball. That is, it is seldom realistic. A more realistic model has it that (16) is replaced by

$$\mathcal{H} : x(t) = \nu(t) \tag{61}$$

$$\mathcal{K} : x(t) = \sum_{r=1}^{R} A_r \cos(2\pi f_c t - 2\pi f_c t_r^{\Delta}) + \nu(t) \tag{62}$$

in which the target is composed[6] of R "reflectors" with strengths $\{A_r\}$ and range-offsets $\{ct_r^{\Delta}/2\}$. We thus write (23) as

$$x^c(t) = p(t) \sum_{r=1}^{R} A_r \cos(\psi_r) + \nu^c(t) \tag{63}$$

$$x^s(t) = p(t) \sum_{r=1}^{R} A_r \sin(\psi_r) + \nu^s(t) \tag{64}$$

in which it is assumed that $t_r^{\Delta} \ll T$ so that the impact on $p(t)$ of ignoring it is small; and all $\{\psi_r\}$ are independent. It is a small step to apply the central limit theorem to write (64) as

$$x^c(t) = A^c p(t) + \nu^c(t) \tag{65}$$

$$x^s(t) = A^s p(t) + \nu^s(t) \tag{66}$$

in which A_c and A_s are zero-mean, independent and Gaussian.

There are many ways to treat this, and the easiest is to note that the analysis in the previous section, in which the target had a constant and known strength A, ended up with a structure that did not depend on A. That is, the sum of the squares of the I & Q matched filters (37) is UMP (uniformly most powerful).

Another approach, which is instructive, is to write vectors x^c & x^s to represent the samples of the I & Q channels, and the vector p containing the samples of $p(t)$, we see that we have a standard Gaussian problem

$$\mathcal{H} : f(x_c, x_s|\mathcal{H}) = \frac{1}{|2\pi \mathbf{R_H}|} e^{-\frac{1}{2}((x^c)^T \mathbf{R}_H^{-1}(x_c) + (x^s)^T \mathbf{R}_H^{-1}(x_s))} \tag{67}$$

$$\mathcal{K} : f(x_c, x_s|\mathcal{K}) = \frac{1}{|2\pi \mathbf{R_K}|} e^{-\frac{1}{2}((x^c)^T \mathbf{R}_K^{-1}(x_c) + (x^s)^T \mathbf{R}_K^{-1}(x_s))} \tag{68}$$

[6]These might be nose, wingtip, engines ... in addition to the specular point, generally any discontinuity and broadside-aspect flat bit reflects energy.

in which

$$\mathbf{R}_H \equiv \sigma^2\mathbf{I} \tag{69}$$

$$\mathbf{R}_K \equiv \sigma^2\mathbf{I} + \sigma_A^2 pp^T \tag{70}$$

and

$$\sigma^2 \equiv \mathcal{E}\{\nu_c^2\} = \mathcal{E}\{\nu_s^2\} \tag{71}$$

$$\sigma_A^2 \equiv \mathcal{E}\{A_c^2\} = \mathcal{E}\{A_s^2\} \tag{72}$$

in which Nyquist-rate sampling is assumed. We thus have the log-likelihood ratio

$$\begin{aligned}\log(L(x_c, x_s)) &= \frac{1}{2}(x^c)^T(\mathbf{R}_H^{-1} - \mathbf{R}_K^{-1})(x_c) + \frac{1}{2}(x^s)^T(\mathbf{R}_H^{-1} - \mathbf{R}_K^{-1})(x_s) \\ &\quad + \log(|\mathbf{R}_H|) - \log(|\mathbf{R}_K|)\end{aligned} \tag{73}$$

Via the Woodbury formula we have

$$\mathbf{R}_K^{-1} = \sigma^{-2}\mathbf{I} - \frac{\sigma^{-4}pp^T}{\sigma_A^{-2} + \sigma^{-2}p^Tp} \tag{74}$$

$$= \mathbf{R}_H^{-1} - \frac{\sigma^{-4}}{\sigma_A^{-2} + \sigma^{-2}p^Tp}pp^T \tag{75}$$

meaning that a sufficient statistic for detection is

$$\lambda(x_c, x_s) = \left(p^T(x_c)\right)^2 + \left(p^T(x_s)\right)^2 \tag{76}$$

meaning that the optimal test is the same as in (37): the sum of the squares of the I & Q channel matched filters – the same result as we would get with the UMP insight.

As for performance, note that with u and v as defined in (34), we have the same $\mathcal{H}$ analysis as (40) leading to (43). Under $\mathcal{K}$ we have – much simpler than the nonfluctuating case – we also have u and v independent and zero-mean Gaussian. That is, the statistical situation under the Swerling I target-present hypothesis $\mathcal{K}$ is the same as under $\mathcal{H}$, the only difference being that the variance is not σ^2 but σ^2+E, in which E is the target energy[7] returned to the receiver. We get

$$\alpha = Pr(T > \tau\sigma^2|\mathcal{H}) \tag{77}$$

$$= e^{-\tau/2} \tag{78}$$

[7]Note that this is the *average* target energy, since the amount of energy returned is a random variable.

and

$$\begin{aligned}\beta &= Pr(T > \tau\sigma^2|\mathcal{K}) && (79)\\ &= \exp\left(\frac{-\sigma^2\tau}{2(\sigma^2+E)}\right) && (80)\\ &= \alpha^{\frac{1}{1+SNR}} && (81)\end{aligned}$$

in which the SNR term is as in (60). We often write

$$\begin{aligned}f(\lambda|\mathcal{H}) &= \frac{1}{\mu}e^{-\lambda/\mu}u(\lambda) && (82)\\ f(\lambda|\mathcal{K}) &= \frac{1}{\mu(1+SNR)}e^{-\lambda/\mu(1+SNR)}u(\lambda) && (83)\end{aligned}$$

as a short form for the Swerling I model.

2.3 The Swerling III Model

The Swerling III model is a modification on the Swerling I idea, in that it is assumed that in (62) one reflector – perhaps the specular – is too large to be lumped in with the others. The upshot is that (66) becomes

$$\begin{aligned}x^c(t) &= (A\cos(\psi)+A^c)p(t) \;+\; \nu^c(t) && (84)\\ x^s(t) &= (A\sin(\psi)+A^s)p(t) \;+\; \nu^s(t) && (85)\end{aligned}$$

in which A_c and A_s are zero-mean, independent and Gaussian, ψ is uniform on $(0, 2\pi)$ and A denotes the strength of the dominant reflector.

Via the UMP insight the test base on the magnitude-square matched filter (37) remains optimal. And since the second and third terms in (85) can be combined, the probability of detection is similar to the non-fluctuating case

$$\beta \;=\; \int_{\sqrt{\tau}\sigma}^{\infty} \frac{r}{\sigma^2+\sigma_t^2} e^{-\frac{r^2+\rho^2}{\sigma^2+\sigma_t^2}} I_0(r\rho/(\sigma^2+\sigma_t^2))dr \qquad (86)$$

in which ρ is as before and σ_t^2 is the twice the variance of A^c and A^s. We approximate

$$I_0(z) \;\approx\; 1 + z^2/4 \qquad (87)$$

and get

$$\begin{aligned}\beta &= \int_{\sqrt{\tau}\sigma}^{\infty} \frac{r}{\sigma^2+\sigma_t^2} e^{-\frac{r^2+\rho^2}{\sigma^2+\sigma_t^2}} \left[1+\frac{1}{4}\left(\frac{r\rho}{\sigma^2+\sigma_t^2}\right)^2\right] dr && (88)\\ &\approx \int_{\sqrt{\tau}\sigma}^{\infty} \frac{r^3\rho^2}{4(\sigma^2+\sigma_t^2)^3} e^{-\frac{r^2+\rho^2}{\sigma^2+\sigma_t^2}} dr && (89)\end{aligned}$$

where the second term is assumed larger than the first[8]. From (89) we can change variables and write

$$f(\lambda|\mathcal{H}) = \frac{1}{\mu}e^{-\lambda/\mu}u(\lambda) \quad (90)$$

$$f(\lambda|\mathcal{K}) = \frac{4\lambda}{(\mu(1+SNR))^2}e^{-2\lambda/\mu(1+SNR)} \quad (91)$$

for Swerling III. Most targets, lore has it, are covered by Swerling I or Swerling III on a single-pulse level. Swerling I is more commonly assumed.

2.4 Swerling II and Swerling IV

Many radar systems operate using multiple pulses, each of length T but separated in time by and inter-pulse interval t_p. The question arises: are the pulses *coherent* (meaning that the "random" phase for each one ψ is the same) or independent? Radar engineers invoke the *coherence time* or *decorrelation time* to discuss this.

Swerling II and Swerling IV are exactly like Swerling I and Swerling III, respectively; except that for Swerling I the pulses are perfectly correlated and for Swerling II they are perfectly independent, with the corresponding relationship between Swerling III and Swerling IV. Clearly for Swerling I and Swerling III the multiple pulses might as well be thought of as a single pulse.

For Swerling II noncoherent integration – that is, summing the λ's from the pulses – is the optimal test statistic. The optimal test in the Swerling IV case is slightly more complicated, but generally noncoherent integration or simple sign detection[9] is used.

In many (or even most) radar situations a target can be considered non-fluctuating over the time period of the pulse-train. Consider a radar with a frequency of 3GHz, a pulse-length of 1μs, a pulse-repetition interval (PRI) or 20μs and 10 pulses: the overall dwell time is approximately 200μs. If the target aspect changes little during this time (as would be expected) the fluctuation would be minor. However, it should be noted that for a Swerling I target (and to a lesser extent a Swerling III target) the lack of fluctuation can actually be a concern, since the target can be in "deep fade" of the sort that communications engineers fear. Fortunately, by varying the center frequency of the pulses in the pulse train an artificially noncoherent target can

[8]This seems to conflict with the Taylor approach. There is admittedly little justification in the Swerling III model.

[9]This means test each pulse separately and count the number of successes.

be manufactured: non-coherent pulse integration (or even decision fusion) should be used for detection.

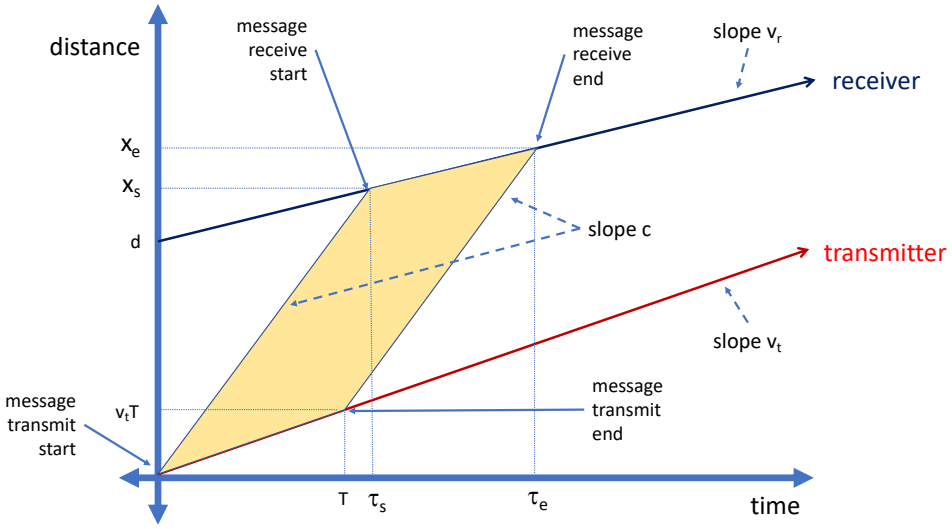

Figure 1: Figure used to explain derivation of Doppler shift for acoustic sources. The horizontal axis is time, the vertical axis is separation between the transmitter and receiver. The transmitter moves at velocity v_t and the receiver v_r; both can be negative, but neither larger than the speed of propagation c. The message duration is T. The goal of the analysis is to compare the apparent pulse time $\tau_e - \tau_s$ as observed by the receiver to T: the ratio is the time dilation factor, and we will soon calculate it to be (108).

3 Doppler Shift

3.1 Narrowband and Slow-Speed Doppler Shift.

This analysis is approximate; we will improve it later. Assume that the relative range-rate of target and transmitter is v; that is, the distance to the target is $d(t) = d_0 + vt$, where the base distance is d_0. The signal that arrives at the target is

$$\begin{aligned} x(t) &= s(t - 2d(t)/c) && (92)\\ &= s(t - 2(d_0 + vt)/c) && (93)\\ &= s(t(1 - 2v/c) - d_0/c) && (94) \end{aligned}$$

which implies that the signal is "stretched" by this time-dilation factor $\alpha = (1 - 2v/c)$. For radar it is reasonable[10] to assume that this is small relative to the pulse bandwidth: $p(t) \approx p(\alpha t)$. However, since generally the carrier frequency f_c is very high we cannot ignore its effect on that; we get,

$$\begin{aligned} x(t) &= \Re\left\{Ap(\alpha t)e^{j2\pi f_c \alpha t}\right\} && (95) \\ &\approx \Re\left\{Ap(t)e^{j2\pi f_c t}e^{j2\pi f_d t}\right\} && (96) \end{aligned}$$

(we are not interested in noise here) in which

$$f_d = \left(\frac{-2v}{c}\right) f_c \qquad (97)$$

is the Doppler shift frequency; $c \approx 3 \times 10^8 m/s$ is the speed of light and f_c is the radar's carrier frequency. Note that v is the *range-rate* of the target to the radar: if the target is "closing" ($v < 0$) then the frequency shift is positive. If $v = 60m/s$ and $f_c = 3GHz$ (a typical S-band[11]), then $f_d = 1200Hz$.

3.2 General Acoustic Doppler Shift

The primary difference between acoustic and electromagnetic Doppler shifts is relativity. Special relativity implies that for electromagnetic the speed of propagation is the same for all frames of reference, transmitter and receiver both. The true Doppler uses the Lorentz contraction factor to calculate the time compression/dilation at the receiver, and is complicated; fortunately the answer turns out to be the same as what is given below, and well approximate by what was just shown, above.

For acoustic sources we do not always have $v \ll c$, since (for example) in air $c \approx 300m/s$, and such speeds are easily within reach. But we also have that the speed of propagation is relative to the medium (assumed at rest), and since both transmitter and receiver can be moving, this can present added difficulties. We will begin a development, and refer to figure 1. The goal of the analysis is to calculate $\tau_e - \tau_s$, since the ratio of that to T is the time dilation factor α.

We have the following equations.

$$\begin{aligned} x_s &= d + v_r \tau_s && (98) \\ x_s &= c\tau_s && (99) \end{aligned}$$

[10] A (very) fast target might be $1.5km/s$; in that case we have $\alpha = 1 \pm 0.00001$.

[11] L-band is 1-2GHz, S-band 2-4GHz, C-band 4-8 GHz ... there are many bands beyond.

$$x_e = x_s + v_r(\tau_e - \tau_s) \tag{100}$$
$$x_e = v_t T + c(\tau_e - T) \tag{101}$$

Combining (98) and (99) we get

$$\tau_s = \frac{d}{c - v_r} \tag{102}$$
$$x_s = \frac{cd}{c - v_r} \tag{103}$$

hence equating the right sides of (100) and (101) we get

$$x_s + v_r(\tau_e - \tau_s) = v_t T + c(\tau_e - T) \tag{104}$$
$$\frac{cd}{c - v_r} + v_r\left(\tau_e - \frac{d}{c - v_r}\right) = v_t T + c(\tau_e - T) \tag{105}$$

or

$$\tau_e = \frac{d - (v_t - c)T}{c - v_r} \tag{106}$$

Combining (102) and (106) we get

$$\tau_e - \tau_s = \frac{c - v_t}{c - v_r} T \tag{107}$$

or

$$\alpha = \frac{1 - v_t/c}{1 - v_r/c} \tag{108}$$

where v_t is positive *toward* the receiver and v_r is positive *away from* the transmitter, as indicated in figure 1. Note that for $|v_r| \ll c$ we use the Taylor approximation to get

$$\alpha \approx (1 - v_t/c)(1 + v_r/c) \tag{109}$$
$$\approx 1 - \frac{v_t - v_r}{c} \tag{110}$$

for the (approximate) *one-way* acoustic Doppler shift.

For the two-way Doppler shift the roles of receiver and transmitter get reversed, for the path from target (now a "transmitter" of its reflected energy) back to the radar receiver (assumed colocated with the original "transmitter"). Consequently the two-way dilation factor is

$$\alpha = \left(\frac{1 - v_t/c}{1 - v_r/c}\right)\left(\frac{1 + v_r/c}{1 + v_t/c}\right) \tag{111}$$

Again using the first-order Taylor expansion we approximate

$$\alpha \approx (1 - v_t/c)(1 + v_r/c)(1 + v_r/c)(1 - v_t/c) \quad (112)$$

$$\approx 1 - 2\frac{v_t - v_r}{c} \quad (113)$$

for the (approximate) *two-way* acoustic Doppler shift. It should be stressed again that (110) and (113) are approximate relations, valid only when $|v_t| \ll c$ and $v_t \ll c$ – more generally (108) holds.

As a final note, it should be observed that the analysis here assumes that during the time of interest (essentially, $T + 2d/c$) the transmitted and receiver can be modeled as having a fixed scalar velocity, as in figure 1. In cases that the full vectors v_r and v_t are not aligned with the vector from transmitter to receiver – and T is large – this may not be true. A visceral example of this is when an ambulance siren is heard passing near: its pitch changes smoothly[12] from elevated in frequency to depressed, as it passes.

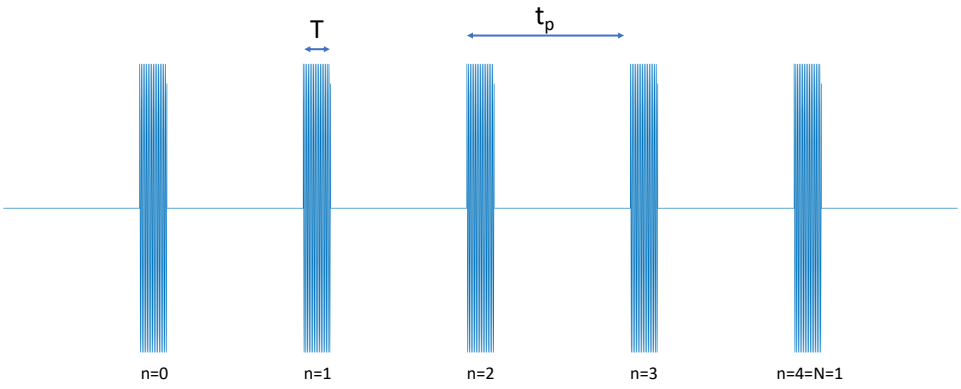

Figure 2: Unnecessary embellishment of what a radar pulse composed of multiple pulses might look like.

3.3 Multiple Pulses

Let us consider the approximate Doppler shift example given previously: for $v = 60m/s$ and $f_c = 3GHz$ we had $f_d = 1200Hz$. Since the Doppler resolution is generally thought to be approximately the reciprocal of the pulse length, a $10\mu s$ pulse (actually rather long for radar) could not reliably observe such a shift, since $1/T = 100kHz$.

[12] In the case that v_t is collinear with the transmitter/receiver vector the pitch switches abruptly. However, the time of switch corresponds to the moment that the receiver is run over by the ambulance, hence this situation is not, strictly speaking, observable.

Now, let us consider the (36) form of (37)

$$\lambda \;=\; \left| \int \tilde{x}(t) p^*(t) dt \right|^2 \tag{114}$$

For a multiple-pulse system (with coherence assumed[13]), we have

$$s(t) \;=\; \Re \left\{ \sum_{n=0}^{N-1} Ap(t - nt_p) e^{j2\pi f_c t} \right\} \tag{115}$$

in which $p(t)$ represents an individual pulse, t_p is the interval separating pulse beginnings, and presumably $t_p \gg T$ in which T is the pulse length – see figure 2. We therefore adapt (36) to this new baseband pulse

$$\sum_{n=0}^{N-1} Ap(t - nt_p) \tag{116}$$

and baseband return

$$\sum_{n=0}^{N-1} Bp(t - nt_p) e^{j2\pi f_d t} + \tilde{\nu}(t) \tag{117}$$

and write

$$\begin{aligned}
\lambda \;&=\; \left| \int_0^{(N-1)t_p+T} \left(\sum_{n=0}^{N-1} Bp(t - nt_p) e^{j2\pi f_d t} + \nu(t) \right) \right. \\
&\qquad \left. \times \left(\sum_{n=0}^{N-1} p^*(t - nt_p) \right) dt \right|^2 \qquad (118) \\
&=\; \left| \sum_{n=0}^{N-1} \int_{nt_p}^{nt_p+T} \Big(Bp(t - nt_p) e^{j2\pi f_d t} + \nu(t) \Big) \right. \\
&\qquad \left. \times \left(p^*(t - nt_p) \right) dt \right|^2 \qquad (119) \\
&=\; \left| \sum_{n=0}^{N-1} \int_0^T B|p(\tau)|^2 e^{j2\pi f_d(\tau + nt_p)} d\tau \right. \\
&\qquad \left. + \int_0^T p^*(\tau) \nu(\tau + nt_p) d\tau \right|^2 \qquad (120) \\
&=\; \left| \sum_{n=0}^{N-1} \left(\int_0^T B|p(\tau)|^2 e^{j2\pi f_d \tau} d\tau \right) e^{j2\pi (f_d t_p) n} \right|^2
\end{aligned}$$

[13]It should be obvious that multi-pulse Doppler is possible with Swerling I & III models; and completely impossible for Swerling II & IV.

$$+ \int_0^T p^*(\tau)\nu(\tau + nt_p)d\tau \Big|^2 \tag{121}$$

$$= \left| \sum_{n=0}^{N-1} ye^{j2\pi(f_d t_p)n} + \nu_n \right|^2 \tag{122}$$

in which the definitions of y and ν_n are obvious. The point is that the Doppler frequency f_d can be inferred from the sequence of single-pulse complex-baseband matched filters via a simple frequency estimation routine. If we return to our example at the beginning of this section, we have $f_d = 1200Hz$. Let us select $t_p = 100\mu s$; then from (122) we have a digital frequency $\omega = 2\pi f_d t_p = 2\pi(0.12)$. With (say) $N = 20$ pulses this is easilty resolvable.

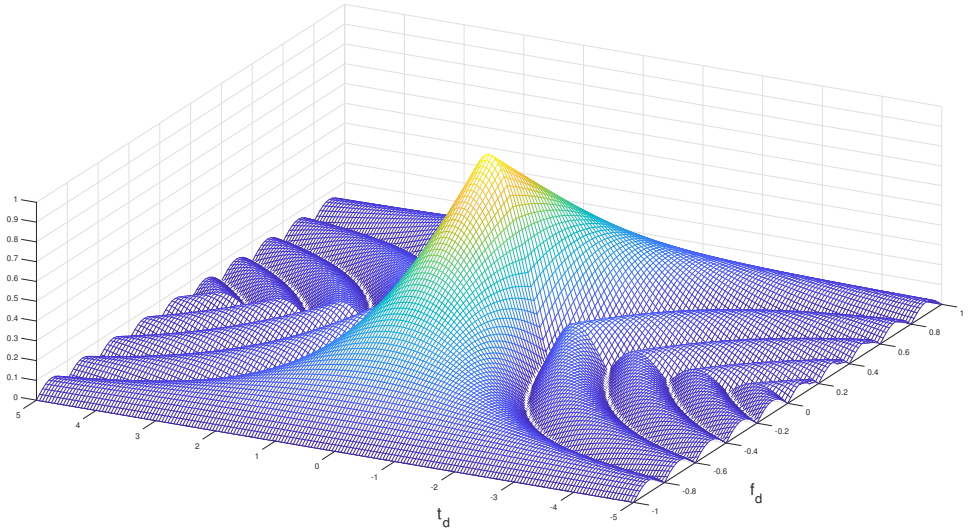

Figure 3: An ambiguity function for the rectangular constant-frequency (CF) pulse case. The time axis extends $\pm T$. The frequency axis is in units of $1/T$.

4 The Ambiguity Function

We have from (36) the matched filter expression

$$\lambda = \left| \int_0^T \tilde{x}^*(t)p(t)dt \right|^2 \tag{123}$$

for the matched filter output. This filter is "matched" to a return exactly aligned in time with its sample, and it assumes no Doppler shift. Let us ignore noise, but allow for mismatch, meaning that we have

$$\tilde{x}(t) \;=\; Ap(t-t_d)e^{j2\pi f_d(t-t_d)} \tag{124}$$

where the delay and Doppler are respectively t_d and f_d.The response clearly depends on the *ambiguity function*

$$\mathcal{A}(t_d,f_d)^2 \;\equiv\; \kappa\left|\int_0^T p(t)p(t-t_d)^* e^{j2\pi f_d t}dt\right|^2 \tag{125}$$

where the normalizer κ scales the maximum value to unity and due to symmetry the sign of t_d does not matter. For a rectangular pulse we have

$$\mathcal{A}(t_d,f_d)^2 \;=\; \kappa\left|\int_0^{T-|t_d|} e^{j2\pi f_d t}dt\right|^2 \tag{126}$$

$$\mathcal{A}(t_d,f_d) \;=\; |(T-|t_p|)\mathrm{sinc}(f_d(T-|t_p|))| \tag{127}$$

This is plotted in figure 3. Note that for this (rectangular and constant-frequency) pulse, with $t_d = 0$ we have the first Doppler "null" at $|f_d| = t/T$ – and greater than than if $t_d \neq 0$. This informs our earlier insight that Doppler sensitivity is inversely proportional to the pulse-length.

The ambiguity function implies the response of a matched filter that is *mis*matched to the truth. That is, it provides insight as to the amount to signal energy that is lost by testing for a target at (t_0, f_0) when the actual target is at (t_1, f_1). How densely must matched filters be placed in delay-Doppler space? Clearly too many matched filters wastes computation; but perhaps worse, it can produce a centroiding issue that a given point target actually appear in more than one (or even many) apparent "hits" – additional logic must be applied to provide a single return with reportable quality. On the other hand, if there are too few matched filters such that that target is missed, that is even worse. Figure 4 shows how the ambiguity function provides guidance: samples should taken densely enough that an acceptable "scalloping loss" (a rule of thumb might be that the ambiguity function be no smaller than 50%) is preserved.

The ambiguity function has some simple properties that are useful to recall

Unimodality. The unique maximum value of the ambiguity function is found at $(t_d, f_d) = (0, 0)$. This is simple to prove via the Schwarz Inequality.

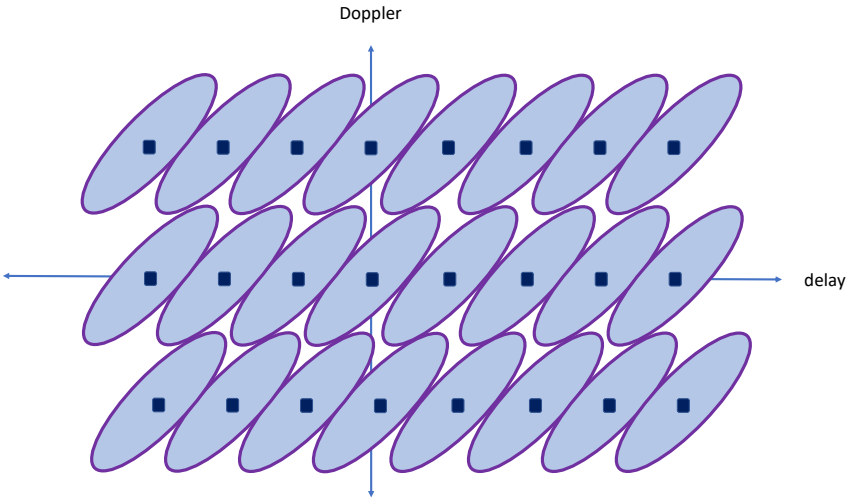

Figure 4: A notional figure indicating how the ambiguity function can suggest matched-filter samples corresponding to *resolution cells*.

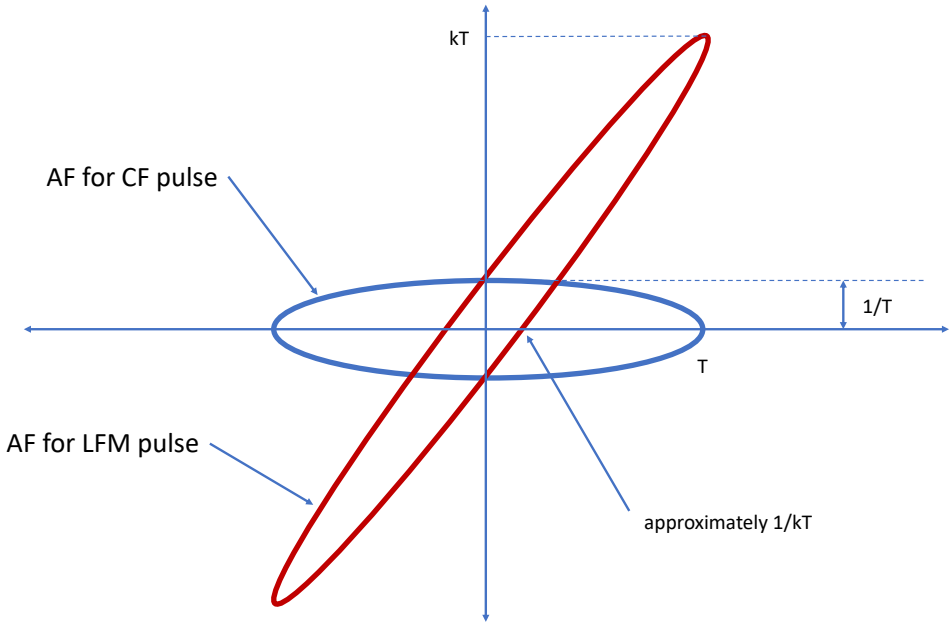

Figure 5: A notional figure indicating the ambiguity function for a rectangular chirp (LFM) waveform.

Unit Volume. Via Parseval it can be shown that we have

$$\int\int \mathcal{A}(t_d, f_d)^2 dt_d df_d \;=\; 1 \tag{128}$$

This has the nice implication that attempting to "push down" a sidelobe by clever waveform design necessarily makes another sidelobe pop up somewhere else.

Chirp. Suppose the baseband pulse $p(t)$ has ambiguity function $\mathcal{A}_p(t_d, f_d)$. Then the chirped pulse $q(t) = p(t)e^{j\pi kt^2}$ has ambiguity function

$$\begin{aligned}
\mathcal{A}_q(t_d, f_d) &= \left| \int_0^T q(t) q(t - t_d)^* e^{j2\pi f_d t} dt \right| && (129)\\
&= \left| \int_0^T p(t) e^{j\pi kt^2} p(t - t_d)^* e^{-j\pi k(t+t_d)^2} e^{j2\pi f_d t} dt \right| && (130)\\
&= \left| e^{-j\pi k t_d^2} \int_0^T p(t) p(t - t_d)^* e^{j2\pi (f_d - k)t} dt \right| && (131)\\
&= \left| \int_0^T p(t) p(t - t_d)^* e^{j2\pi (f_d - kt_d)t} dt \right| && (132)\\
&= \mathcal{A}_p(t_d, f_d - kt_d) && (133)
\end{aligned}$$

The latter property is sketched in figure 5. The original CF pulse ambiguity function is indicated by an ellipse, and the LFM by a tilted ellipse. Assuming the slopes of the CF ellipse near the origin to be zero and of the LFM ellipse to be k, the range resolution is easily seen to be approximately $1/kT$. It turns out that kT is also the bandwidth of the chirp, and the time resolution of any waveform is generally thought to be well approximated by the reciprocal of the bandwidth.

5 Monopulse Radar

The left part of figure 6 sketches a notional radar system. Let us insert some typical numbers: the range of the target is 100km and the beam-width is 10mrad – the latter is about half a degree, which is not so bad although big radars can do better. Simple geometry gives us that that angular accuracy – the arc length of the beam – is 1000m. If we also insert that the radar signal bandwidth is 300MHz, the (approximate!) range accuracy is of the order of 1m. Clearly the cross-range uncertainty is not commensurate with the down-range accuracy.

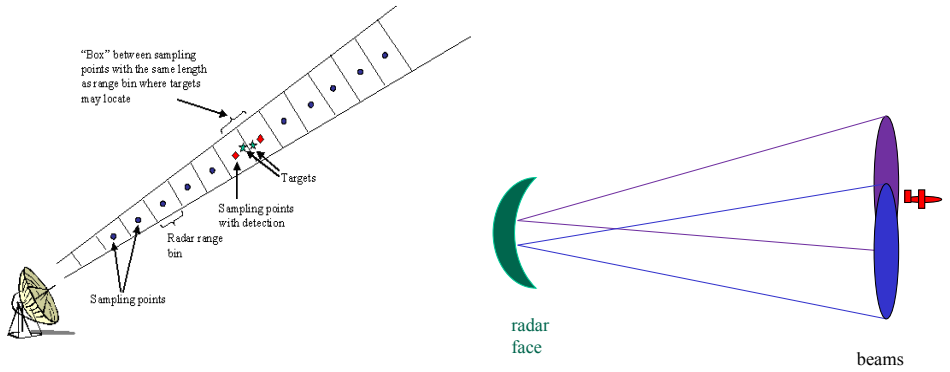

Figure 6: Left: A notional radar without monopulse processing, figure from Blair & Ogle. Right: Monopulse cartoon of the sort that we will consider. Actually there would be four beams to get high resolution in both azimuth and elevation.

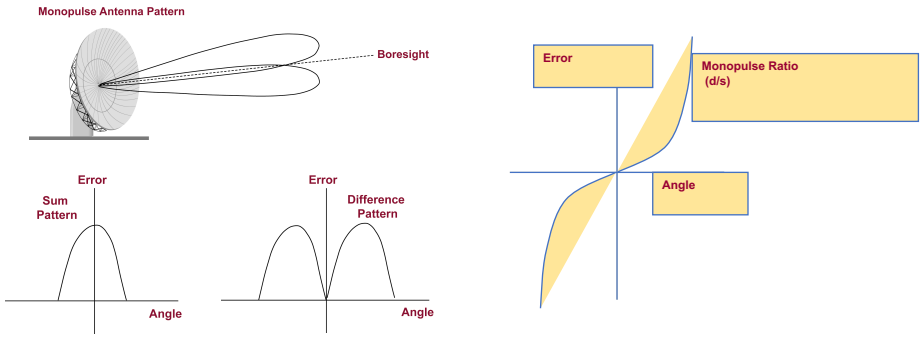

Figure 7: Illustration of construction of the sum and difference channels. Note that close to the boresight the difference channel is approximately linear in the offset from boresight. Figure from Blair & Ogle.

The right part of figure 6 suggests the notion of monopulse radar. One may think of this as each "beam" being comprised of two beams, slightly offset from one another. The usual model for monopulse returns, particularized to one dimension, is

$$s = x + \nu_s \tag{134}$$
$$d = \eta x + \nu_d \tag{135}$$

in which s is the sum channel observation, d is the difference-channel observation, x is a complex Gaussian random variable with mean zero and

variance σ_s^2 that corresponds to the "skin return" from the target, and ν_d and ν_s are complex Gaussian noises each with mean zero and variance σ^2 – all random quantities are independent. The scalar $\eta \in (-1, 1)$ is the *electronic angle* that corresponds to the ratio of the sum of the beam outputs to the difference; again, see figure 7 for what is meant. Note that a Swerling I fluctuation model is here assumed, although it is not necessary.

As a notional example, let us assume the canonical beam pattern is a cosine[14], such that the left and right outputs[15] are

$$y_l = \cos(\beta(\theta - \theta_0))y \tag{136}$$
$$y_r = \cos(\beta(\theta + \theta_0))y \tag{137}$$

in which y is the return from the target, θ is the actual off-boresight angle of the target and θ_0 is the off-boresight steering angle of the two beams – respectively to the left and to the right – and β determines the width of the beam. Then we have

$$\begin{aligned} s &= y_l + y_r && (138)\\ &= [\cos(\beta(\theta - \theta_0)) + \cos(\beta(\theta + \theta_0))]\, y && (139)\\ &= [\cos(\beta\theta)\cos(\beta\theta_0) + \sin(\beta\theta)\sin(\beta\theta_0) \\ &\qquad + \cos(\beta\theta)\cos(\beta\theta_0) - \sin(\beta\theta)\sin(\beta\theta_0)]\, y && (140)\\ &= 2\cos(\beta\theta)\cos(\beta\theta_0)y && (141) \end{aligned}$$

We similarly have

$$\begin{aligned} d &= y_l - y_r && (142)\\ &= 2\sin(\beta\theta)\sin(2\beta\theta_0)y && (143) \end{aligned}$$

Now, consider that we form the *monopulse ratio*

$$\begin{aligned} \hat{\eta} &\equiv \Re\left\{\frac{s^* d}{|s|^2}\right\} && (144)\\ &\approx \tan(\beta\theta_0)\tan(\beta\theta) && (145)\\ &= \eta(\theta) && (146)\\ &\approx [\beta\tan(\beta\theta_0)]\,\theta && (147) \end{aligned}$$

in which the first approximation is that the noises ν_s and ν_d can be neglected, and the second is that the angle θ is small. As a practical note, the cosine-beampattern analysis is just notional: a similar result would be obtained for

[14] Assume that β is small enough that the first zero is beyond the beamwidth.

[15] We here exemplify azimuth with left & right. An analysis for elevation using up and down would be equivalent.

any beampattern that has a reasonable roll-off[16] away from center. As an exercise, the reader is encouraged to examine the raised cosine or Gaussian shape. In (146) the functional notation (that η is actually $\eta(\theta)$) is what practical systems use: they know the sum and difference beampatterns and hence the mapping $\eta(\theta)$: it would make little sense to ignore that, hence of course they don't.

The concept of a sum and difference channel dates to older radar systems in which there were actually four receiver "horns" located slightly offset from the focus of the parabolic antenna. In fact, the original concept for radar was to use a single beam but to examine the target from multiple dwells – to the left, to the right, above and below – and to infer the fine-resolution angle from the differences in received energy. The monopulse system[17] described here is more modern and far more accurate. However, even more modern systems use phased-array radar, in which beams are "steered" based on hundreds of small antenna feeds (actually usually slots in an antenna face). The one sum and two (azimuthal and elevation) difference channels are formed via digital signal processing, and often the two shapes are chosen via Taylor and Bayliss shading, respectively. One might reasonably ask why do this: why not simply infer the target's angle directly from the hundreds of antenna returns. The answer is that the monopulse ratio is close to optimal for angular estimation, and is very fast; nonetheless some moderns radars with good computational abilities do their direction-finding directly from the array element data, using (for example) the MUSIC direction-finding technique.

Let us return to (144) and find the statistics. Conditioned on s we see that d is Gaussian with

$$\mathcal{E}\{d|s\} = \frac{\eta\sigma_s^2}{\sigma_s^2+\sigma^2}s \tag{148}$$

$$\mathcal{V}\{d|s\} = \eta^2\sigma_s^2+\sigma^2-\left(\frac{\eta\sigma_s^2}{\sigma_s^2+\sigma^2}\right)^2(\sigma_s^2+\sigma^2) \tag{149}$$

$$= \sigma^2\left(1+\frac{\eta^2\sigma_s^2}{\sigma_s^2+\sigma^2}\right) \tag{150}$$

[16]Monopulse radar is one situation in which the canonical "brick-wall" filter shape is not a good idea. Actually a triangular beam would be nice in terms of simplicity, as the *relative* strength on the channels is key to angular resolution.

[17]The "original" system with the multiple dwells inferred fine target cross-range position via differences between several radar pulses. When the monopulse system was designed it was so called because it did the same thing with one pulse. However, modern radars use pulse trains to get Doppler – leading to the confusing monicker "multipulse monopulse".

This implies that given s,

$$\hat{\eta} \;=\; \frac{1}{|s|^2}\Re\{s^* d\} \tag{151}$$

is Gaussian as well. Again conditioned on s the ratio $s^* d/|s|^2$ is Gaussian with

$$\mathcal{E}\left\{ \frac{s^* d}{|s|^2} \,\middle|\, s \right\} \;=\; \frac{\eta\sigma_s^2}{\sigma_s^2+\sigma^2} \tag{152}$$

$$\mathcal{V}\left\{ \frac{s^* d}{|s|^2} \,\middle|\, s \right\} \;=\; \frac{\sigma^2}{|s|^2}\left(1+\frac{\eta^2\sigma_s^2}{\sigma_s^2+\sigma^2}\right) \tag{153}$$

It is especially interesting that the conditioning is only on the magnitude-square of the sum-channel return. Defining the sum-channel SNR and the *observed* SNR as

$$\mathcal{S} \;=\; \frac{\sigma_s^2}{\sigma^2} \tag{154}$$

$$\mathcal{S}_o \;=\; \frac{|s|^2}{\sigma^2} \tag{155}$$

we are left with

$$\mathcal{E}\left\{\hat{\eta}\,|\,\mathcal{S}_o\right\} \;=\; \frac{\eta\mathcal{S}}{\mathcal{S}+1} \;\approx\; \eta \tag{156}$$

$$\mathcal{V}\left\{\hat{\eta}\,|\,\mathcal{S}_o\right\} \;=\; \frac{1}{\mathcal{S}_o}\left(1+\frac{\eta^2\mathcal{S}}{\mathcal{S}+1}\right) \;\approx\; \frac{1+\eta^2}{\mathcal{S}_o} \tag{157}$$

where the assumption is that the SNR is relatively large. That is, the *observed* angular accuracy is inversely proportional to the observed SNR, and this is reportable, for example, to a tracker that requires individualized and time-varying measurement accuracy.

6 Constant False-Alarm Rate

6.1 CA-CFAR

We will inform this discussion by the Swerling I example, for which

$$f(\lambda|\mathcal{H}) \;=\; \frac{1}{\mu}e^{-\lambda/\mu}u(\lambda) \tag{158}$$

$$f(\lambda|\mathcal{K}) \;=\; \frac{1}{\mu(1+S)}e^{-\lambda/\mu(1+S)}u(\lambda) \tag{159}$$

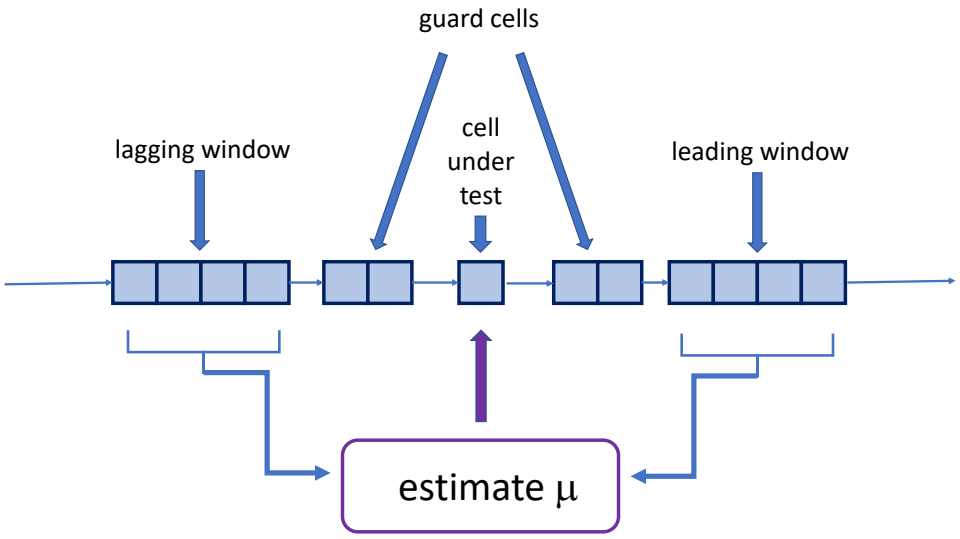

Figure 8: For CFAR operation the estimate of the local noise power is taken from nearby resolution cells. It is often a good idea to exclude nearby "guard band" cells, since they may contain target energy and hence bias the noise-power estimate upwards resulting in target *masking*. In this figure it is implied that the CFAR reference cells are in range only – while this is convenient, it may be useful to use cells nearby in azimuth and / or elevation and / or Doppler. Note that the reference windows should be local to the test cell, since clutter conditions can change greatly, spatially.

in which λ is the test-statistic (sum of squares of I & Q matched-filter outputs), S is the SNR and μ is the noise power. In order to set the threshold for an acceptable false-alarm rate we require knowledge of μ. The CFAR idea is to estimate μ based on other matched filter outputs, as suggested in figure 8.

The *cell-averaging* (CA-CFAR) idea is to use a simple empirical mean to estimate μ – actually this is the optimal estimator based on most reasonable criteria, assuming a homogeneous window. That is, we have

$$\hat{\mu} \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i \;\equiv\; \frac{1}{n} y \tag{160}$$

in which $\{y_i\}_{i=1}^{n}$ are the n reference cells, assumed independent and distributed as exponential with mean μ. We have that the sum y has Gamma

pdf

$$f(y) \;=\; \frac{y^{n-1}}{(n-1)!\mu^n} e^{-y/\mu} u(y) \tag{161}$$

Let us assume that the test cell x has pdf

$$f(x) \;=\; \frac{1}{\lambda\mu} e^{-x/(\mu\lambda)} u(x) \tag{162}$$

where under $\mathcal{H}$ we have $\lambda = 1$ and under $\mathcal{K}$ we have $\lambda = (1+S)$, in which S is the SNR. Then we have

$$\begin{aligned}
Pr(x > \tau y/n) &= \int_0^\infty \int_{\tau y/n}^\infty \frac{1}{\mu\lambda} e^{-\frac{x}{\mu\lambda}} \frac{y^{n-1}}{(n-1)!\mu^n} e^{-\frac{y}{\mu}} u(y) dx dy & (163)\\
&= \int_0^\infty e^{-\frac{\tau y}{n\mu\lambda}} \frac{y^{n-1}}{(n-1)!\mu^n} e^{-\frac{y}{\mu}} u(y) dy & (164)\\
&= \int_0^\infty \frac{y^{n-1}}{(n-1)!\mu^n} e^{-y\left(\frac{\tau}{n\lambda\mu}+\frac{1}{\mu}\right)} u(y) dy & (165)\\
&= \left(\frac{\tau}{n\lambda\mu} + \frac{1}{\mu}\right)^{-n} \mu^{-n} & (166)\\
&= \frac{1}{(1+\frac{\tau}{n\lambda})^n} & (167)
\end{aligned}$$

Note that this does not depend on μ – meaning that this scheme *is* CFAR – and also that we have

$$\lim_{n\to\infty} Pr(x > \tau y/n) \;=\; e^{-\tau/\lambda} \tag{168}$$

which is satisfying. At any rate, we have

$$\begin{aligned}
\alpha &= \frac{1}{(1+\frac{\tau}{n})^n} & (169)\\
\beta &= \frac{1}{\left(1+\frac{\tau}{n(1+S)}\right)^n} & (170)
\end{aligned}$$

In general, $n = 8$ is quite large enough.

6.2 OS-CFAR

A concern with CA-CFAR is that it is sensitive to target *masking* by elements in the reference window that are not homogeneous with the others. This may happen, for instance, when there are closely-spaced targets: it would be a bad system indeed if the best counter-measure were to have more than

one target. Use of an order-statistic (OS-CFAR) approach seems to be considerably more robust.

Suppose we order the elements of the reference window, $\{y_i\}$, from smallest to largest, according to

$$y_{(1)} \ \leq \ y_{(2)} \ \leq \ y_{(3)} \ \leq \ \ldots \ \leq \ y_{(n)} \tag{171}$$

According to Bayes, we have

$$f(\{y_{(i)}\}|\{y_i\})f(\{y_i\}) \ = \ f(\{y_i\}|\{y_{(i)}\})f(\{y_{(i)}\}) \tag{172}$$

Since the first item on the LHS is deterministic, we have

$$\begin{aligned} f(\{y_{(i)}\}) \ &= \ \frac{f(\{y_{(i)}\}|\{y_i\})}{f(\{y_i\}|\{y_{(i)}\})} f(\{y_i\}) && (173) \\ &= \ n! f(\{y_{(i)}\}) && (174) \\ &= \ \frac{n!}{\mu^n} e^{-\frac{1}{\mu}\sum_{i=1}^{n} y_i} && (175) \end{aligned}$$

Suppose we define

$$\delta_i \ \equiv \ \begin{cases} y_{(1)} & i = 1 \\ y_{(i)} - y_{(i-1)} & \text{else} \end{cases} \tag{176}$$

We can re-write (175) as

$$\begin{aligned} f(\{y_i\}) \ &= \ \frac{n!}{\mu^n} e^{-\frac{1}{\mu}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\delta_i\right)} && (177) \\ &= \ \frac{n!}{\mu^n} e^{-\frac{1}{\mu}\sum_{i=1}^{n}(n-i+1)\delta_i} && (178) \end{aligned}$$

This gives the – very surprising! – result that the *differences* $\{\delta_i\}$ are independent and exponentially-distributed, meaning

$$f(\delta_i) \ = \ \frac{n-i+1}{\mu} e^{-\frac{n-i+1}{\mu}\delta_i} \tag{179}$$

This means that if x has an exponential pdf with mean $\mu\lambda$ (parsimonious development as for CA-CFAR), our exceedance probability is

$$Pr(x > \tau y_{(k)}) \ = \ Pr\left(x > \tau\left(\sum_{i=1}^{k}\delta_i\right)\right) \tag{180}$$

$$
\begin{aligned}
&= \int_0^\infty \dots \int_0^\infty e^{-\frac{\tau\left(\sum_{i=1}^k \delta_i\right)}{\mu\lambda}} \left(\frac{\prod_{i=1}^k (n-i+1)}{\mu^k}\right) \\
&\qquad \times e^{-\sum_{i=1}^k \left(\frac{n-i+1}{\mu}\right)\delta_i} d\delta_1 \dots d\delta_k \qquad (181)\\
&= \int_0^\infty \dots \int_0^\infty \left(\frac{\prod_{i=1}^k (n-i+1)}{\mu^k}\right) \\
&\qquad \times e^{-\sum_{i=1}^k \left(\frac{n-i+1}{\mu}+\frac{\tau}{\mu\lambda}\right)\delta_i} d\delta_1 \dots d\delta_k \qquad (182)\\
&= \frac{\left(\frac{\prod_{i=1}^k (n-i+1)}{\mu^k}\right)}{\prod_{i=1}^k \left(\frac{(n-i+1)}{\mu}+\frac{\tau}{\mu\lambda}\right)} \qquad (183)\\
&= \frac{1}{\prod_{i=1}^k \left(1+\frac{\tau}{(n-i+1)\lambda}\right)} \qquad (184)\\
&= \frac{1}{\prod_{i=n-k+1}^n \left(1+\frac{\tau}{i\lambda}\right)} \qquad (185)
\end{aligned}
$$

We thus get

$$
\begin{aligned}
\alpha &= \frac{1}{\prod_{i=n-k+1}^n \left(1+\frac{\tau}{i}\right)} \qquad (186)\\
\beta &= \frac{1}{\prod_{i=n-k+1}^n \left(1+\frac{\tau}{i(1+S)}\right)} \qquad (187)
\end{aligned}
$$

which again are independent of μ and are both nicely computable.

ECE 6124
Signal Detection
Small Random Signals

Peter Willett

Fall 2018

1 The Test

Suppose we have observations

$$x_{mn} \;=\; \nu_{mn} + \theta s_n, \qquad m = 1, 2, \ldots, M \quad \& \quad n = 1, 2, \ldots, N \tag{1}$$

The "noise" ν_{mn} is *iid* with pdf $f(\cdot)$. Notionally (1) models an array detection situation in which the array is *steered* (with appropriate delays) such that the "signal" at each time n is the same at each array element.

The "signal" s_n is <u>not</u> independent, but instead has joint pdf $f_s(\cdot)$. The model is not completely general since the noise could also be dependent; but that the signal $\{s_m\}$ could actually be $\{s_{mn}\}$ is actually equivalent to the case $N = 1$ and M replaced by MN. Note that if one is interested in the time-series detection problem (no array) simply set M to unity.

We have

$$f(x) \;=\; \int \prod_{m=1}^{M} \prod_{n=1}^{N} f(x_{mn} - \theta s_n) f_s(s) ds \tag{2}$$

and hence

$$\frac{d}{d\theta} f(x) \;=\; \int \sum_{m=1}^{M} \sum_{n=1}^{N} P_{mn}(\theta)(-s_n) \dot{f}(x_{mn} - \theta s_n) f_s(s) ds \tag{3}$$

in which

$$P_{mn}(\theta) \;\equiv\; \frac{\prod_{r=1}^{M} \prod_{s=1}^{N} f(x_{rs} - \theta s_r)}{f(x_{mn} - \theta s_n)} \tag{4}$$

If any of the s_n's has a non-zero mean then we are done: the test would be

$$T_{lo}(x) \;=\; \sum_{m=1}^{M}\sum_{n=1}^{N}\mu_n g_{lo}(x_{mn}) \tag{5}$$

which is not so interesting – but what perhaps *is* interesting is that even if only one μ_n is non-zero, that value of n is all that is taken into account in the test.

At any rate, if all the μ_n are zero we need to take another derivative. We get

$$\begin{aligned}&\frac{d^2}{d\theta^2}f(x) \qquad (6)\\ &\quad\int\sum_{m=1}^{M}\sum_{n=1}^{N}P_{mn}(\theta)(s_n^2)\ddot{f}(x_{mn}-\theta s_n)f_s(s)ds\\ &\quad+\int\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{p=1}^{M}\sum_{q=1}^{N}P_{(mn),(pq)}(\theta)s_n s_q\dot{f}(x_{mn}-\theta s_n)\dot{f}(x_{pq}-\theta s_q)f_s(s)ds\end{aligned}$$

in which

$$P_{(mn),(pq)}(\theta) \;\equiv\; \frac{\prod_{r=1}^{M}\prod_{s=1}^{N}f(x_{rs}-\theta s_r)}{f(x_{mn}-\theta s_n)f(x_{pq}-\theta s_q)} \tag{7}$$

Now since the generalized Neyman-Pearson Lemma tells us that our locally-optimal test statistic ought to be

$$T_{lo}(x) \;=\; \frac{\frac{d^2}{d\theta^2}f_\theta(x)|_{\theta=0}}{f_\theta(x)|_{\theta=0}} \tag{8}$$

we get

$$\begin{aligned}T_{lo}(x) \;&=\; \int\sum_{m=1}^{M}\sum_{n=1}^{N}(s_n^2)h_{lo}(x_{mn})f_s(s)ds \qquad (9)\\ &\qquad+\int\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{p=1}^{M}\sum_{q=1}^{N}s_n s_q g_{lo}(x_{mn})g_{lo}(x_{pq})f_s(s)ds\\ &=\; \sum_{m=1}^{M}\sum_{n=1}^{N}h_{lo}(x_{mn}) \qquad (10)\\ &\qquad+\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{p=1}^{M}\sum_{q=1}^{N}r(n-q)g_{lo}(x_{mn})g_{lo}(x_{pq})\end{aligned}$$

in which we have assumed wlog that s_n is wss and unity power, and we have defined

$$h_{lo}(x) \equiv \frac{\ddot{f}(x)}{f(x)} - (g_{lo}(x))^2 \tag{11}$$

In the case that $M = 1$ (no array) we might write this as

$$T_{lo}(x) = \sum_{n=1}^{N} h_{lo}(x_n) + [g_{lo}(x)]^T \mathbf{R} [g_{lo}(x)] \tag{12}$$

and hence comfort ourselves.

For the case that $f(x)$ is Gaussian we have

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{13}$$

$$\dot{f}(x) = \frac{-x}{\sigma^2} f(x) \tag{14}$$

$$\ddot{f}(x) = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2} \right) f(x) \tag{15}$$

which means that

$$h_{lo}(x) = -\frac{1}{\sigma^2} \tag{16}$$

can be ignored and we get our expected quadratic form.

An especially interesting case arises when $f(x)$ is Laplace, the signal is uncorrelated over time, and $M = 2$. We get

$$f(x) = \frac{1}{2\alpha} e^{-|x|/\alpha} \tag{17}$$

$$\dot{f}(x) = -\text{sign}(x) \frac{1}{\alpha^2} f(x) \tag{18}$$

$$\ddot{f}(x) = (1 - 2\delta(x)) f(x) \tag{19}$$

which, allowing ourselves to ignore the $\delta(x)$, implies that

$$T_{lo}(x) = \sum_{n=1}^{N} \text{sign}(x_{1n}) \text{sign}(x_{2n}) \tag{20}$$

known as the *polarity coincidence correlator*, or PCC. The structure extends nicely: for example with $M = 3$ we get

$$T_{lo}(x) = \sum_{n=1}^{N} [\text{sign}(x_{1n}) \text{sign}(x_{2n}) + \text{sign}(x_{1n}) \text{sign}(x_{3n}) + \text{sign}(x_{2n}) \text{sign}(x_{3n})] \tag{21}$$

If the signal is expected to be consistent – at least in sign – across all array elements, this is a pleasing structure,

2 Performance

2.1 Random Signal Efficacy

We resort to the central limit theorem and efficacy, of course. But there is a subtlety. Consider our previous small-signal development, in which the signal strength is θ_l, the test $\delta(x)$ uses n samples, and we have (no loss of generality)

$$n = \kappa(l)^2 l \tag{22}$$

Then we get

$$\alpha = Q\left(\frac{\tau - \mu_n(0)}{\sigma_n(0)}\right) \tag{23}$$

$$\beta = Q\left(\frac{\tau - \mu_n(\theta_l)}{\sigma_n(\theta_l)}\right) \tag{24}$$

and eventually find

$$\beta = Q\left(Q^{-1}(\alpha) - \frac{\mu_n(\theta_l) - \mu_n(0)}{\sigma_n(0)}\right) \tag{25}$$

which we in our previous development approximated as

$$\beta = Q\left(Q^{-1}(\alpha) - \frac{\theta_l \dot{\mu}_n(0)}{\sigma_n(0)}\right) \tag{26}$$

The problem is that

$$\dot{\mu}_n(0) = \frac{d}{d\theta}\left(\int_x \int_s T(x) f(x - \theta s) f_s(s) ds dx\right)_{\theta=0} \tag{27}$$

$$= \left(\int_x \int_s T(x) s \dot{f}(x - \theta s) f_s(s) ds dx\right)_{\theta=0} \tag{28}$$

$$= \left(\int_x T(x) s \dot{f}(x) dx\right)\left(\int_s f_s(s) ds\right) \tag{29}$$

$$= 0 \tag{30}$$

which should be obvious, anyway, from unbiasedness arguments. That means (26) should be replaced by

$$\beta = Q\left(Q^{-1}(\alpha) - \frac{1}{2}\frac{\theta_l^2 \ddot{\mu}_n(0)}{\sigma_n(0)}\right) \tag{31}$$

Operating in parallel to the previous development we have

$$\beta = Q\left(Q^{-1}(\alpha) - \frac{1}{2}\frac{\theta_l^2\sqrt{n}\ddot{\mu}_n(0)}{\sqrt{n}\sigma_n(0)}\right) \tag{32}$$

$$= Q\left(Q^{-1}(\alpha) - \frac{1}{2}\theta_l^2\sqrt{n}\left[\frac{\ddot{\mu}_n(0)}{\sqrt{n}\sigma_n(0)}\right]\right) \tag{33}$$

Assuming that

$$\xi \equiv \lim_{n\to\infty}\frac{1}{4}\left[\frac{\ddot{\mu}_n(0)}{\sqrt{n}\sigma_n(0)}\right]^2 \tag{34}$$

is a limit that converges, we see that in order for the performance of the test to be nontrivial we must have

$$\theta_l = \frac{\gamma}{l^{1/4}} \tag{35}$$

as opposed to the previous (known signal) case in which we had proportionality to the inverse square root. At any rate, we thence have

$$\beta = Q\left(Q^{-1}(\alpha) - c\frac{\gamma^2}{\sqrt{l}}\sqrt{n}\sqrt{\xi}\right) \tag{36}$$

$$= Q\left(Q^{-1}(\alpha) - \gamma^2\kappa\sqrt{\xi}\right) \tag{37}$$

meaning that for two detectors (A & B) to have the same operating point we have

$$ARE_{B,A} = \frac{n_A}{n_B} \tag{38}$$

$$= \frac{\kappa_A^2}{\kappa_B^2} \tag{39}$$

$$= \frac{\xi_B}{\xi_A} \tag{40}$$

(see (37) for the second line and (22) for the third) as we would expect with efficacy. The new pattern for the (slower) way that the signal strength decreases to zero in (35) is interesting and perhaps informing of the renewed difficulty in detecting random signals; but it is perhaps more a mathematical note. However, the new definition of efficacy in (34) is a fact and cannot be ignored in our analysis. The factor of 1/4 in (34) does not contribute much, but as a matter of form we keep it.

2.2 General Array Correlator Efficacy Expressions

The locally-optimal test statistic for our model has been shown to be

$$T(x) = \sum_{i=1}^{n} \left[\sum_{j=1}^{M} h_{lo}(X_{ji}) \right] + \sum_{i=1}^{n} \sum_{m=1}^{n} r(i-m) \left[\sum_{j=1}^{M} g_{lo}(X_{ji}) \right] \left[\sum_{p=1}^{M} g_{lo}(X_{pm}) \right] \tag{41}$$

where

$$h_{lo}(x) = \frac{\ddot{f}(x)}{f(x)} - \left[\frac{\dot{f}(x)}{f(x)} \right]^2 \tag{42}$$

$$g_{lo}(x) = -\frac{\dot{f}(x)}{f(x)} \tag{43}$$

and $\dot{f}(\cdot)$ and $\ddot{f}(\cdot)$ are first two derivatives of $f(\cdot)$. Note that in the case $M = 1$ (a single sensor) this reduces to

$$T(x) = \sum_{i=1}^{n} h_{lo}(X_i) + \sum_{i=1}^{n} \sum_{m=1}^{n} r(i-m) g_{lo}(X_i) g_{lo}(X_m) \tag{44}$$

Computing the performance of this test is not trivial, and previous treatments have dealt only with the white noise case. It will be helpful to derive the efficacy of a generalized correlator

$$T(x) = \sum_{i=1}^{n} \left[\sum_{j=1}^{M} h(X_{ji}) \right] + \sum_{i=1}^{n} \sum_{m=1}^{n} \rho(i-m) \left[\sum_{j=1}^{M} g(X_{ji}) \right] \left[\sum_{p=1}^{M} g(X_{pm}) \right] \tag{45}$$

which has the same structure as the LMO detector. For best performance h and g ought to take on their LO values of (42) and (43) according to the *actual* noise density f; and ρ should be the *actual* signal autocorrelation r. However, these may be unknown, and in order that a robust detector be identified it is important to develop an expression for the performance of an arbitrary statistic having the structure of (45).

We calculate the efficacy of a statistic as defined in (45). Here the actual (white) noise has symmetric density $f(\cdot)$, and the actual signal correlation is $r(k)$. We re-write the statistic as

$$T(x) \;=\; T_1(x) \;+\; T_2(x) \;+\; T_3(x) \tag{46}$$

where

$$T_1(x) \;=\; \sum_{i=1}^{n} \sum_{j=1}^{M} e(X_{ji}) \tag{47}$$

$$T_2(x) \;=\; \sum_{i=1}^{n} \sum_{\substack{m=1 \\ m\neq i}}^{n} \rho(m-i) \sum_{j=1}^{M} g(X_{ji}) \sum_{p=1}^{M} g(X_{pm}) \tag{48}$$

$$T_3(x) \;=\; \sum_{i=1}^{n} \sum_{j=1}^{M} \sum_{\substack{p=1 \\ p\neq j}} g(X_{ji}) g(X_{pm}) \tag{49}$$

and

$$e(x) \;=\; h(x) + g^2(x) \tag{50}$$

Proceeding with the numerator first, we write

$$\frac{\partial^2}{\partial\theta^2} \mathcal{E}_\theta \left[T(x)\right]_{\theta=0} \;=\; \frac{\partial^2}{\partial\theta^2} \left[N_1(\theta)\right]_{\theta=0} \;+\; \frac{\partial^2}{\partial\theta^2} \left[N_2(\theta)\right]_{\theta=0} \;+\; \frac{\partial^2}{\partial\theta^2} \left[N_3(\theta)\right]_{\theta=0} \tag{51}$$

where the N_j are expected values of the components of T, as given below. For the first term we write

$$N_1(\theta) \;=\; \mathcal{E}_\theta \left\{ \sum_{i=1}^{n} \sum_{j=1}^{M} e(X_{ji}) \right\} \tag{52}$$

or

$$N_1(\theta) \;=\; \sum_{i=1}^{n} \sum_{j=1}^{M} \int\int e(x_{ji}) f(x_{ji} - \theta s_i) f_s(s) ds dx_{ji} \tag{53}$$

Taking the derivative we get

$$\frac{\partial^2}{\partial\theta^2} \left[N_1(\theta)\right]_{\theta=0} \;=\; \sum_{i=1}^{n} \sum_{j=1}^{M} \int\int e(x_{ji}) s_i^2 \ddot{f}(x_{ji} - \theta s_i) f_s(s) ds dx_{ji} \tag{54}$$

and evaluating at $\theta = 0$ we write

$$\frac{\partial^2}{\partial\theta^2} \left[N_1(\theta)\right]_{\theta=0} \;=\; nM \int e\ddot{f} \tag{55}$$

where for notational ease we have suppressed the dependence on the variable of integration. Proceeding with the second term we write

$$N_2(\theta) \;=\; \mathcal{E}_\theta \left\{ \sum_{i=1}^{n} \sum_{\substack{m=1 \\ m\neq i}}^{n} \sum_{j=1}^{M} \sum_{p=1}^{M} \rho(m-i) g(X_{ji}) g(X_{pm}) \right\} \tag{56}$$

or

$$N_2(\theta) = \sum_{i=1}^{n}\sum_{\substack{m=1\\m\neq i}}^{n}\sum_{j=1}^{M}\sum_{p=1}^{M}\rho(m-i)\int\int\int g(x_{ji})g(x_{pm}) \times f(x_{ji}-\theta s_i)f(x_{pm}-\theta s_m)f_s(s)dsdx_{ji}dx_{pm} \quad (57)$$

Taking the derivative and evaluating at $\theta = 0$ we get

$$\frac{\partial^2}{\partial\theta^2}\left[N_2(\theta)\right]_{\theta=0} = \sum_{i=1}^{n}\sum_{\substack{m=1\\m\neq i}}^{n}\sum_{j=1}^{M}\sum_{p=1}^{M}\rho(m-i)2\left[\int g\dot{f}\right]^2 r(m-i) \quad (58)$$

or

$$\frac{\partial^2}{\partial\theta^2}\left[N_2(\theta)\right]_{\theta=0} = \sum_{i=1}^{n}\sum_{\substack{m=1\\m\neq i}}^{n}\rho(m-i)r(m-i)2M^2\left[\int g\dot{f}\right]^2 \quad (59)$$

or

$$\frac{\partial^2}{\partial\theta^2}\left[N_2(\theta)\right]_{\theta=0} = 4M^2\sum_{k=1}^{n-1}\rho(k)r(k)(n-k)\left[\int g\dot{f}\right]^2 \quad (60)$$

For the third term we write

$$N_3(\theta) = \mathcal{E}_\theta\left\{\sum_{i=1}^{n}\sum_{j=1}^{M}\sum_{\substack{p=1\\p\neq j}}^{M}g(X_{ji})g(X_{pm})\right\} \quad (61)$$

or

$$N_3(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{M}\sum_{\substack{p=1\\p\neq j}}^{M}\int\int\int g(x_{ji})g(x_{pm}) \times f(x_{ji}-\theta s_i)f(x_{pi}-\theta s_i)f_s(s)dsds_{ji}dx_{pi} \quad (62)$$

Taking the derivative and evaluating at $\theta = 0$ we get

$$\frac{\partial^2}{\partial\theta^2}\left[N_3(\theta)\right]_{\theta=0} = \sum_{i=1}^{n}\sum_{j=1}^{M}\sum_{\substack{p=1\\p\neq j}}^{M}2\left[\int g\dot{f}\right]^2 \quad (63)$$

or

$$\frac{\partial^2}{\partial\theta^2}\left[N_3(\theta)\right]_{\theta=0} = 2nM(M-1)\left[\int g\dot{f}\right]^2 \quad (64)$$

Taking all these terms together we can write

$$\frac{\partial^2}{\partial\theta^2}\mathcal{E}_\theta\left[T(x)\right]_{\theta=0} = \tag{65}$$
$$nM\left[\int e\ddot{f}\right] + 2nM\left[(M-1) + 2M\sum_{k=1}^{n-1} r(k)\rho(k)(1-k/n)\right]\left[\int g\dot{f}\right]^2$$

for the numerator.

Now, examining the denominator, we note that to calculate the variance we need an expression for the test statistic's mean value under the noise-alone hypothesis. Recalling that by assumption f is even and g is odd, we can easily obtain

$$\mathcal{E}_0(T(x)) = nM\int ef \tag{66}$$

We split the second moment into six expressions as

$$\mathcal{E}_0(T^2(x)) = D_{11}(0)+D_{22}(0)+D_{33}(0)+2D_{12}(0)+2D_{13}(0)+2D_{23}(0) \tag{67}$$

The first three of these terms can be written as

$$D_{11}(0) = \mathcal{E}_0\left\{\left(\sum_{i=1}^{n}\sum_{j=1}^{M} e(X_{ji})\right)^2\right\} \tag{68}$$
$$= nM\int e^2 f + (n^2M^2 - nM)\left[\int ef\right]^2 \tag{69}$$

$$D_{22}(0) = \mathcal{E}_0\left\{\left(\sum_{i=1}^{n}\sum_{\substack{m=1\\m\neq i}}^{n}\rho(m-i)\sum_{j=1}^{M} g(X_{ji})\sum_{p=1}^{M} g(X_{pm})\right)^2\right\} \tag{70}$$
$$= 4M^2\sum_{k=1}^{n-1}\rho^2(k)(n-k)\left[\int g^2 f\right]^2 \tag{71}$$

$$D_{33}(0) = \mathcal{E}_0\left\{\left(\sum_{i=1}^{n}\sum_{p=1}^{M}\sum_{\substack{p=1\\p\neq j}}^{M} g(X_{ji})g(X_{pi})\right)^2\right\} \tag{72}$$
$$= 2nM(M-1)\left[\int g^2 f\right]^2 \tag{73}$$

The assumption $\int gf = 0$ implies that all three cross-terms are zero, hence we write

$$D_{12}(0) \;=\; D_{13}(0) \;=\; D_{23}(0) \tag{74}$$

Combining all terms we write

$$\begin{aligned}\mathcal{V}_0\left\{T(x)\right\} \;&=\; nM\left(\int e^2 f - \left[\int ef\right]^2\right) \qquad (75)\\ &\quad + 2nM\left((M-1) + 2M\sum_{k=1}^{n-1}\rho^2(k)(1-k/n)\right)\left[\int g^2 f\right]^2\end{aligned}$$

for the denominator.

Combining the numerator and denominator and taking the limit as $n \to \infty$, we write

$$\xi \;=\; \frac{M}{4}\frac{\left(\int e\ddot{f} \;+\; 2\left[-1 + M\sum_{k=-\infty}^{\infty} r(k)\rho(k)\right]\left[\int g\dot{f}\right]^2\right)^2}{\left(\int e^2 f \;-\; \left[\int ef\right]^2 \;+\; 2\left[-1 + M\sum_{k=-\infty}^{\infty}\rho^2(k)\right]\left[\int g^2 f\right]^2\right)} \tag{76}$$

for the efficacy, and replacing e by $h + g^2$ we get the result

$$\xi \;=\; \frac{M}{4}\frac{\left(\int [h+g^2]\ddot{f} \;+\; 2\left[-1 + M\sum_{k=-\infty}^{\infty} r(k)\rho(k)\right]\left[\int g\dot{f}\right]^2\right)^2}{\left(\int [h+g^2]^2 f \;-\; \left[\int [h+g^2]f\right]^2 \;+\; 2\left[-1 + M\sum_{k=-\infty}^{\infty}\rho^2(k)\right]\left[\int g^2 f\right]^2\right)} \tag{77}$$

We can use Parseval's relation to write

$$\sum_{k=-\infty}^{\infty} r_1(k)r_2(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_1(\omega)\phi_2(\omega)d\omega \tag{78}$$

in which

$$\phi_i(\omega) \;=\; \sum_{k=-\infty}^{\infty} r_i(k)e^{-j\omega k} \tag{79}$$

is the spectrum corresponding to an autocorrelation sequence r_i. We can identify r_1 and r_2 variously as r (the actual correlation) and as ρ (the assumed correlation), and can hence re-write (77) as

$$\begin{aligned}\xi \;&= \qquad (80)\\ &\frac{M}{4}\frac{\left(\int [h+g^2]\ddot{f} \;+\; 2\left[-1 + \frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_r(\omega)\phi_\rho(\omega)d\omega\right]\left[\int g\dot{f}\right]^2\right)^2}{\left(\int [h+g^2]^2 f \;-\; \left[\int [h+g^2]f\right]^2 \;+\; 2\left[-1 + \frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_\rho^2(\omega)d\omega\right]\left[\int g^2 f\right]^2\right)}\end{aligned}$$

It is instructive to note:

- With an accurate estimate of the signal correlation (i.e. $\rho(k) = r(k)$), (77) reduces to:

$$\xi \;=\; \frac{M}{4}\frac{\left(\int[h+g^2]\ddot{f} \;+\; 2\left[-1+\frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_r^2(\omega)d\omega\right]\left[\int g\dot{f}\right]^2\right)^2}{\left(\int[h+g^2]^2 f \;-\; [\int[h+g^2]f]^2 \;+\; 2\left[-1+\frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_r^2(\omega)d\omega\right][\int g^2 f]^2\right)} \tag{81}$$

- With an accurate estimate of the noise density f, we can write $g = g_{lo}$ and $h = h_{lo}$; the efficacy expression (77) thus reduces to:

$$\xi \;=\; \frac{M}{4}\frac{\left(\int\frac{(\ddot{f})^2}{f} \;+\; 2\left[-1+\frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_r(\omega)\phi_\rho(\omega)d\omega\right][I(f)]^2\right)^2}{\left(\int\frac{(\ddot{f})^2}{f} \;+\; 2\left[-1+\frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_\rho^2(\omega)d\omega\right][I(f)]^2\right)} \tag{82}$$

  where $I(f) = \int\frac{(\dot{f})^2}{f}$ is Fisher's information.

- The efficacy of a locally-optimal detector can be written as

$$\xi \;=\; \frac{M}{4}\left(\int\frac{(\ddot{f})^2}{f} \;+\; 2\left[-1+\frac{M}{2\pi}\int_{-\pi}^{\pi}\phi_r^2(\omega)d\omega\right][I(f)]^2\right) \tag{83}$$

In the latter two cases we have used the fact that $\mathcal{E}_0\{h_{lo}+g_{lo}^2\} \equiv 0$. In general $\mathcal{E}_0\{h+g^2\} \neq 0$. As expected, the performance improves with increased correlation between signal samples. It is instructive to observe (for the locally-optimal case) that with $r(k) \equiv 1$, the efficacy is infinite – this is tantamount to the known signal problem, for which the efficacy of (77) is inappropriate and must be modified.

The equations above represent different levels of knowledge about the statistical test. In (83) all parameters are known and the (LO) test is used. In (82) the noise density f is known, but the precise form of the signal correlation structure is not. This is a situation in which it is reasonable to seek a *robust* ρ, the performance of whose corresponding detector is not degraded as the actual correlation (r) varies over its uncertainty class. In (81) the correct signal correlation is known but the noise density is not, and it is reasonable to seek a detector whose nonlinearities g and h engender a similar lack of degradation as the actual noise density f varies.

ECE 6124
Signal Detection
Composite Testing

Peter Willett

Fall 2018

1 The Big Three

1.1 Introduction

As indicated much earlier in these notes, a composite test is one in which either or both of Ω_H or Ω_K is not a singleton – usually it is Ω_K. Unfortunately, many of our practical testing situations are composite. There were several ideas that might be tried, among them

- Attempt to put a prior on θ. When this is reasonable, the likelihood ratio test can be used.
- Look for a uniformly most-powerful (UMP) test.
- When some part of θ relates to the size of the signal, use a *locally-most powerful* approach. Much of our small-signal analysis has investigated such approaches, under the assumption that for large signals practically any reasonable approach works well so focusing on the most difficult cases is the most rewarding strategy.
- Use a maximum likelihood method.

This section of the course will focus on the last. We will discuss the generalized likelihood ratio test (GLRT), the Wald test and the Rao (-score) test.

1.2 The GLRT

The generalized likelihood ratio (GLR) is

$$T_{GLR}(x) \equiv \frac{\max_{\theta\in\Theta_K}\{f(x|\theta)\}}{\max_{\theta\in\Theta_H}\{f(x|\theta)\}} \tag{1}$$

to be compared to a threshold. Earlier in these notes two example GLRTs were given. One was

$$\begin{aligned} \mathcal{H}: &\quad \mathbf{x} = \nu \\ \mathcal{K}: &\quad \mathbf{x} = \mathbf{e}_\theta + \nu \end{aligned} \tag{2}$$

in which $\mathbf{e}_j$ is the j^{th} Cartesian basis vector, and the (obvious) GLRT

$$T(x) = \max_i \{x_i\} \tag{3}$$

was proposed. A second was

$$\mathbf{x} = \mathbf{A}\theta + \nu \tag{4}$$

in which ν is Gaussian with mean zero and covariance $\mathbf{R}_\nu = \sigma^2\mathbf{I}$, and $\mathbf{A}$ is a known skinny matrix – "skinny" means that $\mathbf{A}$ has more columns than rows. This can be formalized as

$$f(\mathbf{x}|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(\mathbf{x}-\mathbf{A}\theta)^T(\mathbf{x}-\mathbf{A}\theta)} \tag{5}$$

with θ an m-vector and where $\Theta_H = \{\mathbf{0}\}$ and $\Theta_K = \{\theta : \theta \neq \mathbf{0}\}$. Provided $m < n$ (n is the length of $\mathbf{x}$) we have the MLE

$$\hat{\theta} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{6}$$

which means that the "signal" in the derived simple hypothesis test is

$$\hat{s} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{7}$$

and hence the GLR (the derived matched filter) is

$$T(\mathbf{x}) = \mathbf{x}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} \tag{8}$$

It is interesting to compare this linear-mode GLRT to another (reasonable) assumption that θ is Gaussian with mean zero and covariance $\mathbf{R}_\theta$. The (most-powerful!) test is now

$$T(\mathbf{x}) = \mathbf{x}^T\left(\mathbf{R}_\nu^{-1} - \left(\mathbf{A}\mathbf{R}_\theta\mathbf{A}^T + \mathbf{R}_\nu\right)^{-1}\right)\mathbf{x} \tag{9}$$

which is not the same. The difference is that there is a prior on θ that biases it toward the origin in (9) but not in (8).

The GLRT is quite general, since maximization does not require that a gradient method or similar be at the heart of finding the MLE – (2) is an obvious example of just that – and maximization can be over both Θ_H and Θ_K. The two tests that follow (Wald and Rao) require that $\Theta_H = \{\theta_0\}$ – that is, that the null hypothesis be a "signal-absent" singleton – and depend loosely on continuity and differentiability of Θ_K. But they are somewhat simpler and can be more implementable.

1.3 The Wald Test

Let us assume that $\Theta_H = \{\theta_0\}$. Suppose we make a small-signal (i.e., Taylor) approximation around an arbitrary value θ_1:

$$\log\left(f(x|\theta_0)\right) \;\approx\; \log\left(f(x|\theta_1)\right) \;+\; (\theta_0 - \theta_1)\left[\frac{d}{d\theta}\log\left(f(x|\theta)\right)\right]_{\theta=\theta_1} \tag{10}$$

This is fine, except that as motivated by the GLRT idea we would like to choose the maximum-likelihood estimate $\theta_1 = \hat{\theta} = \max_{\theta\in\Theta_K}\{f(x|\theta\}$; and for that we usually have

$$\left[\frac{d}{d\theta}\log\left(f(x|\theta)\right)\right]_{\theta=\theta_1} \;=\; 0 \tag{11}$$

since $\hat{\theta}$ is a critical point. As such, it is worth exploring the next Taylor term

$$\begin{aligned}\log\left(f(x|\theta_0)\right) \;\;\approx\;\; & \log\left(f(x|\theta_1)\right) \;+\; (\theta_0 - \theta_1)\left[\frac{d}{d\theta}\log\left(f(x|\theta)\right)\right]_{\theta=\theta_1} \\ & +\; \frac{1}{2}(\theta_0 - \theta_1)^T\left[\frac{d^2}{d\theta}\log\left(f(x|\theta)\right)\right]_{\theta=\theta_1}(\theta_0 - \theta_1)\end{aligned} \tag{12}$$

Now we evaluate this at $\theta_1 = \hat{\theta}$ and get

$$\log\left(\frac{f(x|\hat{\theta})}{f(x|\theta_0)}\right) \;\;=\;\; \log\left(f(x|\hat{\theta})\right) - \log\left(f(x|\theta_0)\right) \tag{13}$$

$$\;\;=\;\; \frac{1}{2}(\theta_0 - \hat{\theta})^T\left[\frac{d^2}{d\theta^2}\log\left(f(x|\theta)\right)\right]_{\theta=\hat{\theta}}(\theta_0 - \hat{\theta}) \tag{14}$$

Now we approximate Fisher's information

$$I(\hat{\theta}) \;\;=\;\; -\,\mathcal{E}\left\{\frac{d^2}{d\theta^2}\log\left(f(x|\theta)\right)\right\} \tag{15}$$

$$\;\;\approx\;\; -\left[\frac{d^2}{d\theta^2}\log\left(f(x|\theta)\right)\right]_{\theta=\hat{\theta}} \tag{16}$$

and write the Wald test statistic as

$$T_{Wald}(x) \;=\; (\hat{\theta} - \theta_0)^T I(\hat{\theta})(\hat{\theta} - \theta_0) \tag{17}$$

The Wald test is eminently justifiable as above. Its (possibly) unattractive feature is that it requires that Fisher's information be calculated for each $\hat{\theta}$ – that is, for each x.

1.4 The Rao Test

It should be understood that subject to regularity conditions – loosely: differentiability and continuity of θ in Θ_K – we have that

$$\hat{\theta} \sim \mathcal{N}(\theta_0, I(\theta_0)^{-1}) \tag{18}$$

assuming the true value of $\theta = \theta_0$ and that $\hat{\theta}$ is its ML estimate. Now consider the "score" (the derivative of the log-likelihood)

$$\frac{d}{d\theta} \log\left(f(x|\theta_1)\right) \approx \frac{d}{d\theta} \log\left(f(x|\theta_0)\right) + (\theta_1 - \theta_0) \left[\frac{d^2}{d\theta^2} \log\left(f(x|\theta)\right)\right]_{\theta=\theta_0} \tag{19}$$

Now if we evaluate (19) at $\theta_1 = \hat{\theta}$ the left side is zero and we have

$$\frac{d}{d\theta} \log\left(f(x|\theta_0)\right) \approx -(\hat{\theta} - \theta_0) \left[\frac{d^2}{d\theta^2} \log\left(f(x|\theta)\right)\right]_{\theta=\theta_0} \tag{20}$$

$$\approx -(\hat{\theta} - \theta_0) I(\theta_0) \tag{21}$$

This implies that

$$\frac{d}{d\theta} \log\left(f(x|\theta_0)\right) \sim \mathcal{N}(0, I(\theta_0)) \tag{22}$$

An obvious goodness-of-fit statistic is therefore

$$T_{Rao}(x) = \left[\frac{d}{d\theta} \log\left(f(x|\theta_0)\right)\right]^T I(\theta_0)^{-1} \left[\frac{d}{d\theta} \log\left(f(x|\theta_0)\right)\right] \tag{23}$$

Two very nice features of the Rao test relative to the Wald test are that no maximization is necessary, nor is there a need to evaluate the Fisher matrix more than once.

However, beyond that, an especially nice Rao test feature is that under $\mathcal{H}$ and asymptotically, the test statistic $T_{Rao}(x)$ is χ^2 with number of degrees of freedom equal to the dimension of θ.

2 Bernoulli Example: Time-Varying Probabilities

2.1 GLRT

Suppose that one models that the probability of success in a series of N trials as being time-varying; that is

$$p_n = Pr(z_n = 1) = 1 - Pr(z_n = 0) \tag{24}$$

where this probability is posited to be known. Does our data $\{z_n\}_{n=1}^N \in \{0,1\}^{\times N}$ fit the model? We have

$$p(z|\theta) \;=\; \prod_{n=1}^{N} \theta_n^{z_n}(1-\theta_n)^{1-z_n} \tag{25}$$

in which $\Theta_H = \{\theta_n = p_n\}$ and $\Theta_K = \{\theta_n \neq p_n\}$.

The generalized likelihood ratio

$$T_{GLR}(z) \;\equiv\; \frac{\max_{\theta\in\Theta_K}\{p(z|\theta))\}}{\max_{\theta\in\Theta_H}\{p(z|\theta)\}} \tag{26}$$

is difficult to beat. Then the denominator is obvious, and

$$T_{GLR}(z) \;=\; \frac{\max_{\{\theta_n\}\neq\{p_n\}}\{\prod_{n=1}^N \theta_n^{z_n}(1-\theta_n)^{1-z_n}\}}{\prod_{n=1}^N p_n^{z_n}(1-p_n)^{1-z_n}} \tag{27}$$

$$\;=\; \left(\prod_{n=1}^{N} p_n^{z_n}(1-p_n)^{1-z_n}\right)^{-1} \tag{28}$$

Taking the logarithm and removing irrelevant constants, we get an equivalent test statistic:

$$T(\mathbf{z}) = \sum_{n=1}^{N} z_n \log\left(\frac{1-p_n}{p_n}\right) \tag{29}$$

which is a weighted sum of the tests' successes. The performance can be predicted via the central limit theorem:

$$\mathcal{E}\{T(\mathbf{z}|H)\} = \sum_{n=1}^{N} \log\left(\frac{1-p_n}{p_n}\right) p_n \tag{30}$$

and variance

$$\mathcal{V}\{T(\mathbf{z})|H\} = \sum_{n=1}^{N} \left(\log\left(\frac{1-p_n}{p_n}\right)\right)^2 p_n(1-p_n) \tag{31}$$

An appropriate test is

$$\text{decision} \;=\; \begin{cases} \text{accept} & T(\mathbf{z}) - \mathcal{E}_0\{T(\mathbf{z})\} \leq \tau \\ \text{reject} & T(\mathbf{z}) - \mathcal{E}_0\{T(\mathbf{z})\} > \tau \end{cases} \tag{32}$$

A satisfactory threshold τ can be calculated as

$$\tau \;=\; Q^{-1}(P_{fa})\sqrt{\sum_{n=1}^{N} \left(\log\left(\frac{1-p_n}{p_n}\right)\right)^2 p_n(1-p_n)} \tag{33}$$

Here $Q(\cdot)$ is the unit-normal tail probability

$$Q(y) \equiv \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{34}$$

$$Pr(x > \tau) = Q\left(\frac{\tau - \mathcal{E}\{x\}}{\sqrt{\mathcal{V}\{x\}}}\right) \tag{35}$$

Assuming the actual probabilities are $\{q_n\}$ instead of $\{p_n\}$, we have

$$P_d = Q\left(\frac{\tau - \sum_{n=1}^N \log\left(\frac{p_n}{1-p_n}\right)(q_n - p_n)}{\sqrt{\sum_{n=1}^N \left(\log\left(\frac{p_n}{1-p_n}\right)\right)^2 q_n(1-q_n)}}\right) \tag{36}$$

to plot our ROC.

For comparison, consider *optimal* (unrealizable) case both $\{q_n\}$ & $\{p_n\}$ known.

$$LLR(\mathbf{z}) = \sum_{n=1}^N z_n \left[\log\left(\frac{q_n}{1-q_n}\right) - \log\left(\frac{p_n}{1-p_n}\right)\right] \tag{37}$$

Example results are shown in figure 1.

Suppose we modified the test such that we required that the alternative hypothesis was that we could only do better than $\{p_n\}$. That would mean that under the alternative hypothesis $\theta_n > p_n$. As usual we write

$$f(z|\theta) = \prod_{n=1}^N \theta_n^{z_n}(1-\theta_n)^{1-z_n} \tag{38}$$

and

$$\Theta_H = \{\theta_n = p_n\} \tag{39}$$

$$\Theta_K = \{\theta_n > p_n\} \tag{40}$$

The GLRT requires

$$\max_{\theta\in\Theta_K}\left\{\prod_{n=1}^N \theta_n^{z_n}(1-\theta_n)^{1-z_n}\right\} = \prod_{z_n=1}(1)^{z_n}\prod_{z_n=0}(1-p_n)^{1-z_n} \tag{41}$$

$$= \prod_{n=1}^N (1-p_n)^{1-z_n} \tag{42}$$

The GLR is thus

$$L(z) = \frac{\prod_{n=1}^N (1-p_n)^{1-z_n}}{\prod_{n=1}^N p_n^{z_n}(1-p_n)^{1-z_n}} \tag{43}$$

$$= \prod_{n=1}^N (1/p_n)^{z_n} \tag{44}$$

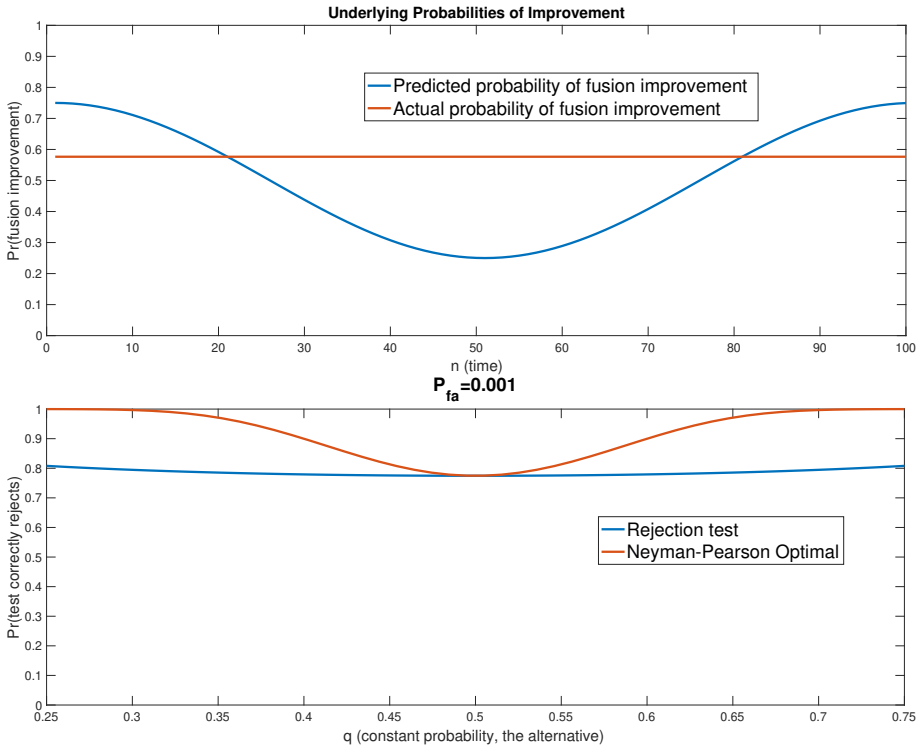

Figure 1: Test for a constant q_n versus a time-varying p_n. Upper plot shows p_n and a particular q_n. Lower plot shows P_d vs level of q_n. The GLRT does acceptably.

An equivalent test statistic is

$$T(z) \;=\; \sum_{n=1}^{N} z_n \log(1/p_n) \tag{45}$$

The mean under H is

$$\mathcal{E}\{T(z)|H\} \;=\; \sum_{n=1}^{N} p_n \log(1/p_n) \tag{46}$$

with corresponding variance

$$\mathcal{V}\{T(z)|H\} \;=\; \sum_{n=1}^{N} p_n(1-p_n)\left(\log(1/p_n)\right)^2 \tag{47}$$

An example of performance and comparison the the "freer" test presented earlier are given in figure 2.

2.2 The Wald Test

Similar to the GLRT, if $\Theta_K = \{\theta_n \neq p_n\}$ then

$$\text{argmax}_{\theta \in \Theta_K} = \{z_n\} \tag{48}$$

Now

$$\begin{aligned} I(\theta) &= \mathcal{E}\left\{ \frac{d^2}{d\theta^2} \log(p(z|\theta)) \right\} && (49) \\ &= \mathcal{E}\left\{ \frac{d^2}{d\theta^2} \sum_{n=1}^{N} [z_n \log(\theta_n) + (1 - z_n) \log(1 - \theta_n)] \right\} && (50) \\ &= \begin{pmatrix} \frac{1}{\theta_1(1-\theta_1)} & 0 & \dots & 0 \\ 0 & \frac{1}{\theta_2(1-\theta_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\theta_N(1-\theta_N)} \end{pmatrix} && (51) \end{aligned}$$

Clearly this presents problems because this is infinite at (48). Put another way, there is no Wald test in this case.

2.3 The Rao Test

From (51) we have that

$$I(\theta_0) = \begin{pmatrix} \frac{1}{p_1(1-p_1)} & 0 & \dots & 0 \\ 0 & \frac{1}{p_2(1-p_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{p_N(1-p_N)} \end{pmatrix} \tag{52}$$

We also have

$$\begin{aligned} \frac{d}{d\theta} \log(p(z|\theta))\Big|_{\theta=\theta_0} &= \frac{d}{d\theta} \sum_{n=1}^{N} [z_n \log(\theta_n) + (1 - z_n) \log(1 - \theta_n)]\Big|_{\theta_n=p_n} && (53) \\ &= \sum_{n=1}^{N} \left[\frac{z_n}{p_n} - \frac{1 - z_n}{1 - p_n} \right] && (54) \end{aligned}$$

These, with (23), tell us that

$$T_{Rao}(z) = \sum_{n=1}^{N} \begin{cases} \frac{p_n}{1-p_n} & z_n = 0 \\ \frac{1-p_n}{p_n} & z_n = 1 \end{cases} \tag{55}$$

It is instructive (but comforting) to note that $\mathcal{E}\{T_{Rao}(z)\} = N$, as predicted.

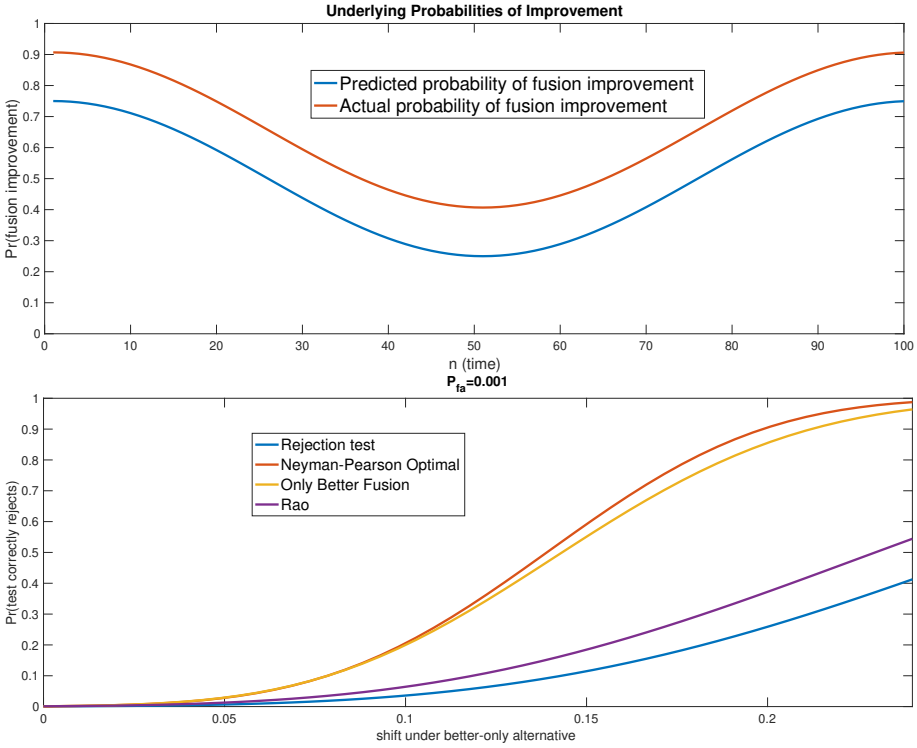

Figure 2: Test $q_n = p_n$ plus a shift. Upper plot shows p_n and a particular q_n. Lower plot shows P_d vs level of q_n. The GLR test is not so good, and the alternative GLR that tests only for an increase in $\{p_n\}$ is more reasonable. The Rao test is better than the one-sided GLRT in this case; but that neither is as good as the alternative GLR test is perhaps a little unfair, since that test uses information not available to them.