$$C36\_40 \qquad C36\_41.$$

ACKNOWLEDGMENT

All the computations were done with the MAGMA system [4].

REFERENCES

[1] C. Bachoc and P. Gaborit, "Designs and self-dual codes with long shadows," *J. Combin. Theory Ser. A*, vol. 105, no. 1, pp. 15–34, 2004.

[2] R. T. Bilous, "Enumeration of the binary self-dual codes of length 34," *J. Combin. Math. Combin. Comput*, vol. 59, pp. 173–211, 2006.

[3] R. T. Bilous and G. H. J. van Rees, "An enumeration of self-dual codes of length 32," *Des. Codes, Cryptogr.*, vol. 26, no. 1-3, pp. 61–86, 2002.

[4] W. Bosma and J. J. Cannon, *Handbook of Magma Functions*, 2.9 ed. Sydney, Australia: Univ. Sydney, 1995 [Online]. Available: http://magma.maths.usyd.edu.au/magma/htmlhelp/MAGMA.htm

[5] J. H. Conway and V. S. Pless, "On the enumeration of self-dual codes," *J. Combin. Theory Ser. A*, vol. 28, pp. 26–53, 1980.

[6] J. H. Conway and N. J. A. Sloane, "A new upper bound on the minimal distance of self-dual codes," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1319–1333, Nov. 1990.

[7] H. Chen, R. Cramer, S. Goldwasse, R. de Haan, and V. Vaikuntanathan, "Secure computation from random error correcting codes," in *Proc. Eurocrypt*, Barcelona, Spain, 2007, pp. 291–310.

[8] S. T. Dougherty, T. A. Gulliver, and M. Harada, "Extremal binary self-dual codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 2036–2047, Nov. 1997.

[9] P. Gaborit, "A bound for certains-extremal lattices and codes," *Arch. Math (Basel)*, vol. 89, no. 2, pp. 143–151, 2007.

[10] M. Harada, "New extremal self-dual codes of lengths 36 and 38," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2541–2543, Nov. 1999.

[11] S. Houghten, C. Lam, L. Thiel, and J. Parker, "The extended quadratic residue code is the only $(48, 24, 12)$ self-dual doubly-even code," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 53–59, Jan. 2003.

[12] J.-L. Kim, "New extremal self-dual codes of lengths 36,38, and 58," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1575–1580, May 2001.

[13] S. Han and J. L. Kim, "Upper bound for the length of s-extremal codes over F2, F4 and F2+uf2," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 418–422, Jan. 2008.

[14] G. Nebe, E. Rains, and N. J. A. Sloane, *Slef-Dual Codes and Invariant Theory*. Berlin, Germany: Springer-Verlag, 2006.

[15] V. Pless, *Introduction to the Theory of Error Correcting Codes*, 3rd ed. New York: Wiley, 1998.

[16] V. Pless, "A classification of self-orthogonal codes over GF(2)," *Discr. Math.*, vol. 3, pp. 209–246, 1972.

[17] V. Pless and N. J. A. Sloane, "On the classification and enumeration of self-dual codes," *J. Combin. Theory Ser. A*, vol. A18, pp. 313–335, 1975.

[18] E. M. Rains and N. J. A. Sloane, "Self-dual codes," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier, 1998, pp. 177–294.

[19] S. T. Dougherty, S. Mesnager, and P. Solé, Secret Sharing Schemes Based on Self-Dual Codes, preprint.

A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data

Liam Paninski

***Abstract*—How many independent samples N do we need from a distribution p to decide that p is ϵ-distant from uniform in an L_1 sense, $\sum_{i=1}^{m} |p(i) - 1/m| > \epsilon$? (Here m is the number of bins on which the distribution is supported, and is assumed known *a priori*.) Somewhat surprisingly, we only need $N\epsilon^2 \gg m^{1/2}$ to make this decision reliably (this condition is both sufficient and necessary). The test for uniformity introduced here is based on the number of observed "coincidences" (samples that fall into the same bin), the mean and variance of which may be computed explicitly for the uniform distribution and bounded nonparametrically for any distribution that is known to be ϵ-distant from uniform. Some connections to the classical birthday problem are noted.**

***Index Terms*—Convex bounds, hypothesis testing, minimax.**

I. INTRODUCTION

We look at a rather basic problem: how many independent and identically distributed (i.i.d.) samples N are required to decide that a discrete distribution p, supported on m points, is nonuniform in an L_1 sense?

Manuscript received June 12, 2006; revised June 19, 2008. Current version published September 17, 2008. This work was supported in part by an NSF CAREER award.

The author is with the Department of Statistics, Columbia University, New York, NY 10027 USA (e-mail: liam@stat.columbia.edu).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2008.928987

More precisely, how large must the sample size N be so that we may test between the null hypothesis

$$H_0 : p_i \equiv 1/m$$

and the nonparametric alternative

$$H_A : \sum_{i=1}^{m} |p(i) - 1/m| > \epsilon$$

with error approaching zero? We will be interested in the sparse case $m \gg N$, where the classical chi-square theory does not apply.

This question has seen a great deal of analysis in both the computer science [1], [2] and statistics [8] literature; in particular, there are obvious connections to the "birthday problem" [5], [7] and related techniques for entropy estimation [12]. In fact, our analysis makes essential use of a version of the so-called "birthday inequality," which states that coincident birthdays are least likely when birthdays are uniformly distributed [3], [6], [11]. The symmetry of the uniform distribution plays a key role here.

It turns out that the uniformity testing problem is easy, in the sense that we may reliably detect departures from uniformity with many fewer samples N than bins m. In fact, it turns out that the condition $N\epsilon^2 m^{-1/2} \to \infty$ guarantees the consistency of a fairly simple test based on the number of "coincidences," samples that fall into the same bin. Thus, for fixed ϵ, we only really need about $N \gg \sqrt{m}$ samples. This is similar in spirit to the recent observation that estimating the entropy of discrete distributions is easy [13] (in that case, $N = cm$ for any $c > 0$ suffices, and hence, by a subsequence argument in fact slightly fewer than $\sim m$ samples are required to estimate the entropy on m bins). Thus, it is much easier to test whether a distribution is uniform than to actually estimate the full distribution (this requires $N \gg m$ samples, as is intuitively clear and as can be made rigorous by a variety of methods, e.g., [4], [14]).

In addition, we prove a lower bound implying that N must grow at least as quickly as $\epsilon^{-2} m^{1/2}$ to guarantee the consistency of *any* test (not just the coincidence-based test introduced here); with fewer samples, any test will fail to detect the nonuniformity of at least one distribution in the alternate class H_A.

II. Upper Bound

Our uniformity test will be based on "coincidences," that is, bins i for which more than one sample is observed. Alternatively, we may look at K_1, the number of bins into which just one sample has fallen; for N fixed, K_1 is clearly directly related to the negative number of coincidences. The basic idea, as in the birthday inequality, is that deviations from uniformity necessarily lead to an increase in the expected number of coincidences, or equivalently, a decrease in $E(K_1)$.

To see this, we may directly write out the expectation of K_1 under a given p, using linearity of expectation

$$E_p(K_1) = \sum_{i=1}^{m} \binom{N}{1} p_i (1-p_i)^{N-1}.$$

In the uniform case, $p_i \equiv 1/m$ and

$$E_u(K_1) = N \left(\frac{m-1}{m}\right)^{N-1}. \tag{1}$$

Now we will compare these two expectations by computing the difference

$$E_u(K_1) - E_p(K_1) = N \left(\frac{m-1}{m}\right)^{N-1} \sum_{i=1}^{m} p_i \left[1 - \left(\frac{m}{m-1}(1-p_i)\right)^{N-1}\right].$$

After some approximations and an application of Jensen's inequality, we have the following key lower bound on $E(K_1)$ in terms of the distance from uniformity ϵ:

Lemma 1:

$$E_u(K_1) - E_p(K_1) \geq \frac{N^2\epsilon^2}{m}[1 + O(N/m)], \quad \forall\, p \in H_A.$$

(A technical note: as noted above, we restrict our attention to the "sparse" regime $N = o(m)$, where direct estimation of the underlying distribution p is not feasible [4], [14].)

Proof: Making the abbreviation

$$f(p_i) = p_i \left[1 - \left(\frac{m}{m-1}(1-p_i)\right)^{N-1}\right]$$

we have

$$E_u(K_1) - E_p(K_1) = N \left(\frac{m-1}{m}\right)^{N-1} \sum_{i=1}^{m} f(p_i). \tag{2}$$

The function $f(x)$ has a fairly simple form: $f(0) = 0, f(1/m) = 0$, $f(x) < 0$ for $0 < x < 1/m$, $f(x)$ is monotonically increasing for $x > 1/m$, and $f(x) \to x$ as x becomes large. However, $f(x)$ is not convex. To develop a lower bound on $E_u(K_1) - E_p(K_1)$, we develop a convex lower bound on $f(x)$, valid for all $x \in [0,1]$ when $N \leq m$

$$f(x) \geq g(|x - 1/m|) + f'(1/m)(x - 1/m)$$

with

$$g(z) = \begin{cases} f(z+1/m) - f'(1/m)z, & z \in [0, 1/N - 1/m] \\ f(1/N) + (z + 1/m - 1/N) - f'(1/m)z, & \text{o.w.} \end{cases}$$

This lower bound on $f(\cdot)$ looks more complicated than it is: for values of x close to $1/m$, where $f(x)$ is convex, we have simply reflected $f(x)$ about the point $1/m$ and added a line in order that the reflected function is smooth. For $x > 1/N$, we have replaced f with a linear lower bound of slope 1 (the limiting slope for $f(x)$ for large x). Here the point $x = 1/N$ is chosen as the solution of the equation

$$f'(x) = 1, \quad 1/m < x < 1;$$

this solution exists uniquely when $m > N$. The derivative $f'(1/m)$ is easily computed as

$$f'(1/m) = (N-1)/(m-1)$$

and similarly, we may directly compute $f'(1/N) = 1$. The key is that $g(|z|)$ is convex, symmetric, and strictly increasing in its argument z; see Fig. 1 for an illustration.

Now we subtract off the line and then apply Jensen. First, we have, for any constant c

$$\begin{aligned} \sum_i [f(p_i) - c(p_i - 1/m)] &= \sum_i f(p_i) - \sum_i c(p_i - 1/m) \\ &= \sum_i f(p_i) - c\left[\left(\sum_i p_i\right) - 1\right] \\ &= \sum_i f(p_i); \end{aligned}$$

in particular, we have that

$$\begin{aligned} \sum_i f(p_i) &\geq \sum_i \left[g(|p_i - 1/m|) + f'(1/m)(p_i - 1/m)\right] \\ &= \sum_i g(|p_i - 1/m|). \end{aligned}$$

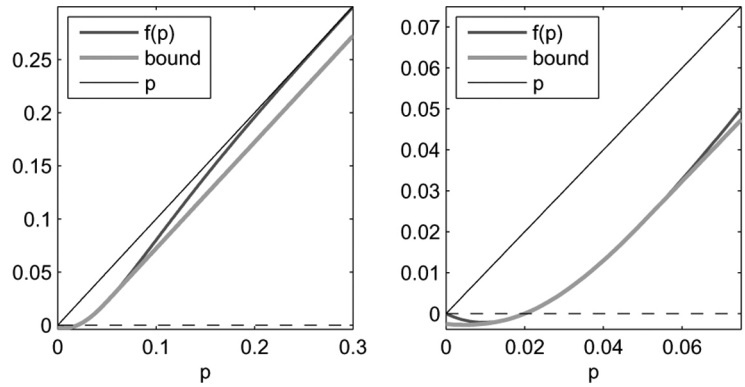

Fig. 1. Illustration of the convex lower bound on the function $f(p)$ in (2). The right panel is just a zoomed-in version of the left panel. $N = 20$; $m = 50$.

Now Jensen implies

$$\frac{1}{m}\sum_i g(|p_i - 1/m|) \geq g\left(\frac{1}{m}\sum_i |p_i - 1/m|\right) \geq g(\epsilon/m)$$

where the last inequality is by the fact that g is increasing and $p \in H_A$. Thus, we find that

$$\sum_i f(p_i) \geq m g(\epsilon/m)$$

and therefore

$$E_u(K_1) - E_p(K_1) \geq N m \left(\frac{m-1}{m}\right)^{N-1} g(\epsilon/m).$$

Now we need to look at $g(\cdot)$. Near 0, $g(\cdot)$ behaves like a quadratic matched to f at the point $1/m$

$$g(z) = \frac{A}{2}z^2 + o(z^2), \quad z \to 0$$

with

$$\begin{aligned} A &= \left.\frac{\delta^2 f(x)}{\delta x^2}\right|_{x=1/m} \\ &= \left(\frac{m}{m-1}\right)^{N-1}\Big[2(N-1)(1-x)^{N-2} \\ &\qquad - (N-1)(N-2)x(1-x)^{N-3}\Big]_{x=1/m} \\ &= 2N + O(N^2/m). \end{aligned}$$

Thus, we have

$$\begin{aligned} E_u(K_1) - E_p(K_1) &\geq N m \left(\frac{m-1}{m}\right)^{N-1} \\ &\quad \times \left([N + O(N^2/m)]\frac{\epsilon^2}{m^2} + o\left(\frac{\epsilon^2}{m^2}\right)\right) \\ &= \frac{N^2\epsilon^2}{m}[1 + O(N/m)], \end{aligned}$$

which completes the proof. □

On the other hand, we may bound the variance of K_1 under p as follows.

Lemma 2:

$$Var_p(K_1) \leq E_u(K_1) - E_p(K_1) + O(N^2/m).$$

Proof: It is not difficult to compute $Var_p(K_1)$ exactly

$$\begin{aligned} &Var_p(K_1) \\ &\quad = E_p(K_1) - E_p(K_1)^2 + N(N-1)\sum_{i\neq j} p_i p_j (1 - p_i - p_j)^{N-2}. \end{aligned}$$

However, we found it inconvenient to bound this formula directly. Instead, we use the Efron–Stein inequality [17]

$$Var(S) \leq \frac{1}{2}E\sum_{j=1}^{N}(S - S^{(i)})^2$$

where S is an arbitrary function of N independent random variables (RVs) x_i and

$$S^{(i)} = S(x_1, x_2, \ldots, x_i', \ldots, x_N)$$

denotes S computed with x_i' substituted for x_i, where x_i' is an i.i.d. copy of x_i. We will apply this inequality to $S = K_1$, with x_i the independent samples from p.

Since we are dealing with i.i.d. samples here, by symmetry we may write

$$\begin{aligned} &\frac{1}{2}E\sum_{j=1}^{N}(S - S^{(i)})^2 \\ &\quad = \frac{N}{2}E_{\{x_i\}_{1\leq i\leq N-1}\sim p}\Bigg[\sum_{i\leq i,j\leq m} p_i p_j \Big(1(n_i = 0 \cap n_j > 0) \\ &\qquad\qquad + 1(n_j = 0 \cap n_i > 0)\Big)\Bigg] \\ &\quad = N\sum_{i,j} p_i p_j P_{\{x_i\}_{1\leq i\leq N-1}\sim p}(n_i = 0 \cap n_j > 0) \\ &\quad = N\sum_{i,j} p_i p_j (1-p_i)^{N-1}\left(1 - \left(1 - \frac{p_j}{1-p_i}\right)^{N-1}\right) \\ &\quad = N\sum_{i,j} p_i p_j \left((1-p_i)^{N-1} - (1-p_i-p_j)^{N-1}\right) \\ &\quad \leq N\sum_{j=1}^{m} p_j \left(1 - (1-p_j)^{N-1}\right) \\ &\quad = E_u(K_1) - E_p(K_1) + N\left(1 - \left(\frac{m-1}{m}\right)^{N-1}\right) \\ &\quad = E_u(K_1) - E_p(K_1) + O(N^2/m). \end{aligned}$$

(Here n_i denotes the number of samples observed to have fallen in bin i after $N-1$ samples have been drawn, the second-to-last equality follows from (2), and the inequality uses the fact that $(1-y)^n - (1-y-x)^n$ is a decreasing function of y for $n > 1$, $x \in [0,1]$, and $0 < y < 1 - x$.) □

Now we may construct our test for H_0 versus H_A: we reject H_0 if

$$T \equiv E_u(K_1) - K_1 = N\left(\frac{m-1}{m}\right)^{N-1} - K_1 > T_\alpha$$

for some threshold T_α.

Theorem 3: The size of this test is

$$P_u(T \geq T_\alpha) = O\left(\frac{N^2}{mT_\alpha^2}\right).$$

The power is greater than

$$P_p(T \geq T_\alpha) \geq 1 - \frac{E_u(K_1) - E_p(K_1) + O(N^2/m)}{(E_u(K_1) - E_p(K_1) - T_\alpha)^2}$$

uniformly over all alternatives $p \in H_A$. If

$$N^2\epsilon^4/m \to \infty$$

then the threshold T_α may be chosen so that the size tends to zero and the power to one, uniformly over all $p \in H_A$ (i.e., this condition is sufficient for the test to be uniformly consistent). For example

$$T_\alpha = N^2\epsilon^2/2m$$

suffices.

We should note that the above bounds are based on a simple application of Chebyshev's inequality and therefore are by no means guaranteed to be tight.

Proof: We have that $E_u(T) = 0$ and $V_u(T) = O(N^2/m)$ (by Lemma 2), and therefore, by Chebyshev, the size is bounded by

$$P_u(T \geq T_\alpha) = O\left(\frac{N^2}{mT_\alpha^2}\right).$$

For the power, we have that

$$\begin{aligned} P_p(T < T_\alpha) &= P_p\left(T - E_p(T) < T_\alpha - E_p(T)\right) \\ &\leq \frac{E_p(T) + O(N^2/m)}{(E_p(T) - T_\alpha)^2} \end{aligned}$$

again by Lemma 2.

Now, by Lemma 1, it is clear that for the size to tend to zero and the power to tend to one, it is sufficient that the "z-score"

$$\frac{N^2\epsilon^2/m}{(N^2/m + N^2\epsilon^2/m)^{1/2}} = \left(\frac{N^2\epsilon^4/m}{1+\epsilon^2}\right)^{1/2}$$

tends to infinity. Since $0 < \epsilon \leq 2$, and $\epsilon^4/(1+\epsilon^2) \sim \epsilon^4$ for $\epsilon \in [0, 2]$, the proof is complete. □

III. Lower Bound

The preceding theorem states that $N^2\epsilon^4/m \to \infty$ is a sufficient condition for the existence of a uniformly consistent test of H_0 versus H_A. The following result is a converse.

Theorem 4: If $N^2\epsilon^4 < m \log 5$, then no test reliably distinguishes H_0 from H_A; more precisely, for any test with critical region B and size bounded away from one, the minimum power

$$\inf_{p \in H_A} \int_B p(x)dx$$

remains bounded away from one.

Proof: It is a well-known [9], [10], [16] consequence of the classical Neyman–Pearson theory that no uniformly consistent test exists if the L_1 distance

$$\left\| u(\vec{x}) - \int_{q \in H_A} q(\vec{x})d\mu(q) \right\|_1$$

is bounded away from 2 for any $\mu \in \mathcal{P}(H_A)$, with $\mathcal{P}(H_A)$ the class of all probability measures on H_A, and $\|\cdot\|_1$ denoting the L_1 norm on the sample space $\vec{x} \in \{1, \ldots, m\}^N$ equipped with the counting measure.

We develop this bound for one particular tractable mixing measure μ. (We make no claims that this measure will lead to optimal bounds.) Assume that m is even. (An obvious modification applies if m is odd.) We choose q randomly according to the following distribution $\mu(q)$: choose $m/2$ independent Bernoulli RVs $z_j \in \{-1, 1\}$ (i.e., z samples uniformly from the corners of the $m/2$-dimensional hypercube). Given $\{z_j\}$, set

$$q(i) = \begin{cases} (1 + \epsilon z_{i/2})/m, & i \text{ even} \\ (1 - \epsilon z_{(i+1)/2})/m, & i \text{ odd.} \end{cases}$$

Such a q will be a probability measure satisfying the equality $\|u - q\|_1 = \epsilon$ (and, therefore, lie on the boundary of the alternate hypothesis class H_A) with probability one, assuming $\epsilon \leq 1$. We let $Q(\vec{x}) = \int q(\vec{x})d\mu(q)$ denote the resulting probability measure. Similar mixing measures have appeared, e.g., in [16]; this mixture of indistinguishable distributions technique is a fundamental idea in the minimax density estimation literature.

To compute the corresponding bound, we use the elegant (albeit somewhat involved) method outlined in Pollard's "Asymptopia" minimax notes [15].

1) First, we substitute a more manageable L_2 bound for the L_1 norm

$$\|Q - u\|_1 \leq \|Q - u\|_2.$$

2) Next, we write out the likelihood ratio

$$Q = 2^{-m/2} \sum_{z \in \{-1,1\}^{m/2}} Q_z$$

with

$$\frac{dQ_z}{du}(\vec{x}) = \prod_{j=1}^{N}(1 + G(x_j, z))$$

where $G(x_j, z) = \epsilon z_{j/2}$ or $-\epsilon z_{(j+1)/2}$, depending as j is even or odd, respectively. Note that

$$E_u G(x_j, z) = 0$$

for all j, z. Define

$$\begin{aligned} \Delta(\vec{x}) &\equiv \frac{dQ}{du}(\vec{x}) \\ &= 2^{-m/2} \sum_z \left(1 + \sum_{j=1}^{N} G(x_j, z) \right. \\ &\qquad \left. + \sum_{j>j'} G(x_j, z)G(x_{j'}, z) + \cdots\right) \end{aligned}$$

the sums ending with the N-fold product.

3) Now we expand the L_2 norm

$$\begin{aligned} \|Q - u\|_2^2 &= E_u(\Delta - 1)^2 \\ (\Delta - 1)^2 &= 2^{-m} \sum_{z,z'} \left(1 + \sum_j G(x_j, z) \right. \\ &\qquad \left. + \sum_{j>j'} G(x_j, z)G(x_{j'}, z') + \cdots\right). \end{aligned}$$

Because of the independence of x_j and z, and the fact that G has zero mean, we may cancel all of the terms that are not products of the form

$$H_j(z,z') \equiv E_u G(x_j,z)G(x_j,z') = \frac{2\epsilon^2}{m}\sum_{i=1}^{m/2} v(z_i,z_i')$$

with $v(z_i,z_i')=1$ if $z_i=z_i'$ and $v(z_i,z_i')=-1$ otherwise. So we have

$$\begin{aligned} E_u(\Delta-1)^2 &= 2^{-m}\sum_{z,z'}\left(\sum_j H_j(z,z') \right. \\ &\qquad \left. + \sum_{j>j'} H_j(z,z')H_{j'}(z,z') + \cdots\right) \\ &= 2^{-m}\sum_{z,z'}\prod_j (1+H_j(z,z')) - 1. \end{aligned}$$

4) The above term may be regarded as an average over two i.i.d. RVs z and z'

$$E_u(\Delta-1)^2 = E_{z,z'}\prod_j (1+H_j(z,z')) - 1.$$

5) Now we use $\log(1+t)\le t$

$$\prod_j (1+H_j(z,z')) \le \exp\left(\sum_j H_j(z,z')\right).$$

6) Finally, we compute

$$\begin{aligned} &E_{z,z'}\exp\left(\sum_j H_j(z,z')\right) \\ &\qquad = \left(\frac{1}{2}\exp\left(\frac{2N\epsilon^2}{m}\right) + \frac{1}{2}\exp\left(-\frac{2N\epsilon^2}{m}\right)\right)^{m/2} \end{aligned}$$

and use the bound

$$\frac{1}{2}\left(\exp(u)+\exp(-u)\right) \le \exp\left(\frac{u^2}{2}\right)$$

to obtain

$$E_{z,z'}\exp\left(\sum_j H_j(z,z')\right) \le \exp\left(\frac{N^2\epsilon^4}{m}\right).$$

Putting everything together

$$\|Q-u\|_1 \le \left(\exp\left(\frac{N^2\epsilon^4}{m}\right)-1\right)^{1/2}.$$

Thus, if $N^2 m^{-1}\epsilon^4$ is not sufficiently large, then $\|Q-u\|_1$ is bounded away from 2, and no uniformly consistent test exists. □

An Alternate Lower Bound

The preceding result provides a quantitative, nonasymptotic lower bound on the error probability, but the bound is loose and the result becomes useless for a fixed ϵ if N^2/m becomes too large. It is worth deriving a simpler, asymptotic result to handle this case of large but bounded N^2/m.

Theorem 5: If N^2/m remains bounded, then no test reliably distinguishes H_0 from H_A.

Proof: The proof here is much more direct. We write out the ratio of marginal likelihoods, using the same uniform-hypercube mixture prior on H_A as above. Letting n_i denote the number of samples observed to have fallen into the ith bin, we have

$$\begin{aligned} \frac{L(\vec{n}|H_A)}{L(\vec{n}|H_0)} &= E_{\vec{z}}\prod_{i=2,4,\ldots,m}(1-z_{i/2}\epsilon)^{n_{i-1}}(1+z_{i/2}\epsilon)^{n_i} \\ &= \prod_{i=2,4,\ldots m} E(1-z_{i/2}\epsilon)^{n_{i-1}}(1+z_{i/2}\epsilon)^{n_i} \\ &= \prod_{i=2,4,\ldots m}\left((1+\epsilon)^{n_{i-1}}(1-\epsilon)^{n_i} \right. \\ &\qquad \left. + (1+\epsilon)^{n_i}(1-\epsilon)^{n_{i-1}}\right)/2 \\ &= \prod_{i=2,4,\ldots m}(1-\epsilon^2)^{m_i}\left((1-\epsilon)^{d_i}+(1+\epsilon)^{d_i}\right)/2 \\ &= \prod_{i=2,4,\ldots m}(1-\epsilon^2)^{m_i}\left(1+\binom{d_i}{2}\epsilon^2+\binom{d_i}{4}\epsilon^4+\cdots\right) \end{aligned}$$

where we have abbreviated $m_i = \min(n_i, n_{i-1})$ and $d_i = |n_i - n_{i-1}|$, and used the independence of z_j. (We interpret $\binom{d_i}{k}$ as 0 whenever $d_i < k$.)

Now note that the above multiplicands are greater than one only if $d_i \ge 2$, and less than one only if $m_i \ge 1$. And, since the number of "two-bin coincidences"—pairs of bins into which two or more samples have fallen—is bounded in probability if $N = O(\sqrt{m})$, the likelihood ratio is bounded in probability as well, implying that the error probability of any test is bounded away from zero, and the proof is complete. Finally, it is worth noting that the expected numbers of the events $(m_i=1, d_i=0)$ and $(m_i=0, d_i=2)$ scale together, leading (after an expansion of the logarithm and a cancellation of the ϵ^2 terms) to exactly the $N^2\epsilon^4/m$ scaling we observed previously. □

REFERENCES

[1] T. Batu, "Testing Properties of Distributions," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 2001.

[2] M. Bellare and T. Kohno, "Hash function balance and its impact on birthday attacks," in *Proc. EUROCRYPT*, Interlaken, switzerland, May 2004, pp. 401–418.

[3] D. Bloom, "A birthday problem," *Amer. Math. Monthly*, vol. 80, pp. 1141–1142, 1973.

[4] D. Braess and H. Dette, "The asymptotic minimax risk for the estimation of constrained binomial and multinomial probabilities," *Sankhya*, vol. 66, pp. 707–732, 2004.

[5] M. Camarri and J. Pitman, "Limit distributions and random trees derived from the birthday problem with unequal probabilities," *Electron. J. Probab.*, vol. 5, pp. 1–18, 2000.

[6] M. Clevenson and W. Watkins, "Majorization and the birthday inequality," *Math. Mag.*, vol. 64, pp. 183–188, 1991.

[7] D. A. Gupta, "The matching, birthday and the strong birthday problem: A contemporary review," *J. Statist. Plann. Inference*, vol. 130, pp. 377–389, 2005.

[8] P. Diaconis and F. Mosteller, "Methods for studying coincidences," *J. Amer. Statist. Assoc.*, vol. 84, pp. 853–861, 1989.

[9] D. Donoho and R. Liu, "Geometrizing rates of convergence," *Ann. Statist.*, vol. 19, pp. 633–701, 1991.

[10] L. LeCam, *Asymptotic Methods in Statistical Decision Theory*. New York: Springer, 1986.

[11] A. Munford, "A note on the uniformity assumption in the birthday problem," *Amer. Statistician*, vol. 31, pp. 119–119, 1977.

[12] I. Nemenman, W. Bialek, and R. de Ruyter van Steveninck, "Entropy and information in neural spike trains: Progress on the sampling problem," *Phys. Rev. E*, vol. 69, pp. 056111–056111, 2004.

[13] L. Paninski, "Estimating entropy on m bins given fewer than m samples," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2200–2203, Sep. 2004.

[14] L. Paninski, "Variational minimax estimation of discrete distributions under KL loss," *Adva. Neural Inf. Process. Syst.*, vol. 17, pp. 1033–1040, 2005.

[15] D. Pollard, Asymptopia 2003 [Online]. Available: www.stat.yale.edu/~pollard
[16] Y. Ritov and P. Bickel, "Achieving information bounds in non- and semi-parametric models," *Ann. Statist.*, vol. 18, pp. 925–938, 1990.
[17] J. Steele, "An Efron-Stein inequality for nonsymmetric statistics," *Ann. Statist.*, vol. 14, pp. 753–758, 1986.

Bayesian Analysis of Interference Cancellation for Alamouti Multiplexing

Songsri Sirianunpiboon, A. Robert Calderbank, *Fellow, IEEE*, and Stephen D. Howard

***Abstract*—Space–time codes built out of Alamouti components have been adopted in wireless standards such as UMTS, IEEE 802.11n, and IEEE 802.16, where they facilitate higher data rates through multiplexing of parallel data streams and the addition of two or more antennas at the receiver that perform interference cancellation. This correspondence provides new theoretical insight into different algorithms for interference cancellation through a Bayesian analysis that expresses performance as a function of signal-to-noise ratio (SNR) in terms of the "angles" between different space–time coded data streams.**

***Index Terms*—Alamouti code, Bayesian analysis, decoding algorithms, interference cancellation, Quaternion arithmetic, space-time block codes.**

I. Introduction

Information-theoretic analysis by Foschini [1] and by Telatar [2] shows that multiple antennas at the transmitter and receiver enable high rate wireless communication. Space–time codes, introduced by Tarokh *et al.* [3], improve the reliability of communication over fading channels by correlating signals across different transmit antennas. The most famous space–time block code (STBC) was discovered by Alamouti [4] and the reason for broad commercial interest in this code is that both coherent and noncoherent detection are remarkably simple. It is possible to separate the data streams transmitted from the two antennas using only linear processing at the receiver. This means that the end-to-end complexity of signal processing is essentially the same as single-antenna systems.

The Alamouti code also facilitates higher data rates through multiplexing of parallel data streams and the addition of a second antenna at the receiver that performs interference cancellation. Data rates of 4 bits/s/Hz have been demonstrated for several wireless channels including UMTS, GSM EDGE, IEEE 802.11n, and IEEE 802.16 (see [5]). In each case, the algebraic structure of the space–time block code makes it possible to implement end-to-end receiver functionality without going beyond the capabilities of digital signal processors (DSPs) used in second-generation cellular technology. Our Bayesian analysis of interference cancellation provides new theoretical insight and is able to predict performance of different detection algorithms as a function of signal-to-noise ratio (SNR). For transmission schemes involving multiplexing a number of independent Alamouti coded transmissions we find for a given overall SNR, the performance is dominated by the channel volume which depends on the complete channel only through the "angles" between the component Alamouti channels.

Manuscript received October 12, 2006; revised April 11, 2008. Current version published September 17, 2008. The work of S. Sirianunpiboon was supported by the Australian Defence Science and Technology Organization (DSTO) long range research fellowship. The work of A. R. Calderbank was supported in part by the National Science Foundation under Grant 0701226 and by the Air Force Office of Scientific Research under MURI Grant AFOSR-FA9550–05–1–0443. The material in this correspondence was presented in part at IEEE Information Theory Workshop, Bergen, Norway, July 2007.

S. Sirianunpiboon and S. D. Howard are with the Defence Science and Technology Organization, PO Box 1500, Edinburgh 5111, Australia (e-mail: songsri.sirianunpiboon@dsto.defence.gov.au; stephen.howard@dsto.defence.gov.au).

A. R. Calderbank is with Electrical Engineering and Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: calderbk@math.princeton.edu).

Communicated by H. Boche, Associate Editor for Communications.

Color versions of Figures 1–3 in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.929012

II. Interference Cancellation of Multiple Alamouti Schemes

The encoding rule for the Alamouti space–time block code is described by a 2×2 matrix

$$(c_1, c_2) \rightarrow \begin{pmatrix} c_1 & c_2 \\ -c_2^* & c_1^* \end{pmatrix} \tag{1}$$

where the columns represent different time slots, the rows represent different antennas, and the entries are the symbols to be transmitted. The signals (r_1, r_2) received over two consecutive time slots are given by

$$\begin{pmatrix} r_1 \\ -r_2^* \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \tag{2}$$

where h_1, h_2 are the path gains from the two transmit antennas to the mobile, and the noise samples n_1, n_2 are independent samples of a zero-mean complex Gaussian random variable with zero mean and covariance $2\sigma^2$. The structure of the Alamouti code induces a structure on the channel. For any complex channel vector (h_1, h_2) the induced channel matrix is

$$(h_1, h_2) \rightarrow H = \begin{pmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{pmatrix}. \tag{3}$$

The space of matrices $\begin{pmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{pmatrix}$ with (h_1, h_2) varying over $\mathbb{C}^2$ form a representation of the *quaternions*.

Consider two co-channel users, each using the Alamouti code. Let $\boldsymbol{c} = (c_1, c_2)^{\mathrm{T}}$ and $\boldsymbol{s} = (s_1, s_2)^{\mathrm{T}}$ be the codewords transmitted by the first and second users, respectively; $\boldsymbol{r}_1 = (r_{11}, -r_{12}^*)^{\mathrm{T}}$ and $\boldsymbol{r}_2 = (r_{21}, -r_{22}^*)^{\mathrm{T}}$ are the received signal vectors, where the components of $\boldsymbol{r}_i$ are the signals received at the antenna i over two consecutive symbols periods. We have

$$\begin{aligned} \boldsymbol{r}_1 &= H_1 \boldsymbol{c} + G_1 \boldsymbol{s} + \boldsymbol{n}_1 \\ \boldsymbol{r}_2 &= H_2 \boldsymbol{c} + G_2 \boldsymbol{s} + \boldsymbol{n}_2 \end{aligned} \tag{4}$$

where H_1 and H_2 are the channel matrices from the first user to the first and second receive antennas, respectively, and the matrices G_1 and G_2 are the channel matrices from the second user to the first and second receive antennas, respectively. The vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are complex Gaussian random variables with zero mean and covariance $2\sigma^2 I_2$, where I_n is $n \times n$ identity matrix.

Rewrite (4) as

$$\boldsymbol{r} = H\boldsymbol{c} + G\boldsymbol{s} + \boldsymbol{n} \tag{5}$$

where

$$\boldsymbol{r} = (\boldsymbol{r}_1, \boldsymbol{r}_2)^{\mathrm{T}}, H = (H_1, H_2)^{\mathrm{T}} \quad \text{and} \quad G = (G_1, G_2)^{\mathrm{T}}.$$