IEEE Information Theory Society Newsletter

Vol. 67, No. 2, June 2017 EDITOR: Michael Langberg, Ph.D. ISSN 1059-2362

President's Column

Rüdiger Urbanke

Some of you might have noticed a monthly email from ITSOC in your mail box with the table of contents of the upcoming IT Transactions (if not, check you spam filter and make sure you paid your dues!). This was suggested by Prakash Narayan, our Editor-in-Chief. Thanks also to Elza Erkip, our First Vice President, for her support, and to Anand Sarwate, the Chair of the Online Committee, for posting this information on our web page. We hope that you like the idea. The early feedback has been very positive. But we can do even better. Dave Forney has suggested that we bring the format into the 21st century and perhaps combine with regular communication to our members. After playing around with various HTML versions and mailers all I can say: Who knew that spamming could be so complicated! Please keep the comments coming.

On a related note, if you are looking for papers from our community make sure that once in a while you check out Xplore. And if you are writing a paper ensure that you also include the reference to the final version of the paper. This is a small step for any author, but has potentially a large impact for our society since it will ensure that our click rates adequately reflect our work and contributions.

As part of the Shannon Centennial Celebration, the IEEE Information Theory Society launched last year a pilot project to create short videos. The objective was to showcase intriguing results from Information Theory (in a broad sense) to a large non-expert audience. Two videos, one on Network Coding and one on Space Time codes, have almost been completed. We have decided to extend this project and we are now looking for volunteers to help define and create the next three videos! If there is a concept or an idea, rooted in information theory, that you think the whole world should know about, this is your chance. A call for video proposal has been released on the ITSOC website with more information. We look forward to your creative ideas and suggestions. http://www.itsoc.org/news-events/recent-news/call-for-short-video-proposal

Talking about movies. Mark Levinson, the director of the Shannon movie, has started editing the material from the February shoot and is making plans for additional shooting of flashback scenes and interviews. He has also begun speaking to a composer and graphics people. The role of Shannon has already been cast—sorry, you can stop practicing juggling. But if you have ideas about how best to explain central concepts of information theory to a large audience, let us know.

Unfortunately, I also have to bring you very sad news. Mary Elizabeth (Betty) Moore Shannon, the remarkable wife of Claude, has died. You can find a reprint of the obituary that has appeared in the Boston Globe and online in the New York Times in this Newsletter.

Besides the Shannon movie, a Shannon biography, authored by Jimmy Soni and Rob Goodman, has just been published. You can order it at https://www.amazon.com/Mind-Play-Shannon-Invented-Information/dp/1476766681

Read it, and if you enjoy it, recommend it to your friends, like it, and up-vote it. An enthusiastic support by our community will help push the book onto bestseller lists, boosting our outreach efforts.

And if you want to do more, follow Christina Fragouli's suggestion. Start your Christmas shopping early this year and get a few copies for your loved ones or anybody who is digitally challenged but not entirely hopeless. Want more ideas? How about donating a copy to your local library or placing a copy in the waiting room of a train station or dentist office. You never know where the next Claude will pass by, looking for inspiration.

(continued on page 10)

From the Editor

Michael Langberg

Dear colleagues,

This issue opens with an intriguing survey article by Adam Smith, "Information, Privacy and Stability in Adaptive Data Analysis," addressing the subtleties in data analytics when collected data is reused. The survey connects the study of adaptive data analytics with certain information measures and measures of distributional stability. The latter, termed "differential privacy," has seen a significant amount of interest lately and has recently awarded Adam Smith (together with Cynthia Dwork, Frank McSherry, and Kobbi Nissim) with the prestigious Gödel Prize on their outstanding work introducing differential privacy. Many thanks to Adam Smith for his generous willingness and significant efforts in preparing the survey.

The issue continues with a number of announcements, regular columns, and reports. We start by congratulating our fellow colleagues for their outstanding research accomplishments and service recognized by our community. Specifically, congratulations to Ido Tal and Alexander Vardy on their 2017 IEEE Communications Society and Information Theory Society award winning paper "List Decoding of Polar Codes," and to the members of our society that have recently been named Distinguished Lecturers. We continue with a special "Behind the scenes Q&A" with the authors of the upcoming Claude Shannon biography. Thanks to authors Jimmy Soni and Rob Goodman for preparing the Q&A and for pursuing this impressive and important project.

We conclude with Tony Ephremides's Historian's column; our "Students' Corner'' column, by Mine Alsan and Basak Güler, presenting the recent initiatives of the Information Theory Student Subcommittee; and a note by Emanuele Viterbo and Elza Erkip, "Thinking about organizing an ITW or ISIT?," outlining the ins and outs of organizing a workshop or conference in our society. Thanks to the contributors for their efforts.

Continuing our remembrance and honoring of Sol Golomb, an extraordinary scholar and long time newsletter contributor, the fourth and final collection of Sol's earlier newsletter puzzle columns appears in this issue. This last collection includes solutions to the second collection of puzzles published in the December 2016 issue of the newsletter.

continued on page 15

IEEE Information Theory Society Newsletter

IEEE Information Theory Society Newsletter (USPS 360-350) is published quarterly by the Information Theory Society of the Institute of Electrical and Electronics Engineers, Inc.

Headquarters: 3 Park Avenue, 17th Floor, New York, NY 10016-5997.

Cost is $1.00 per member per year (included in Society fee) for each member of the Information Theory Society. Printed in the U.S.A. Periodicals postage paid at New York, NY and at additional mailing offices.

Postmaster: Send address changes to IEEE Information Theory Society Newsletter, IEEE, 445 Hoes Lane, Piscataway, NJ 08854.

Table of Contents

Information, Privacy and Stability in Adaptive Data Analysis

Adam Smith*

Abstract

Traditional statistical theory assumes that the analysis to be performed on a given data set is selected independently of the data themselves. This assumption breaks downs when data are re-used across analyses and the analysis to be performed at a given stage depends on the results of earlier stages. Such dependency can arise when the same data are used by several scientific studies, or when a single analysis consists of multiple stages.

How can we draw statistically valid conclusions when data are re-used? This is the focus of a recent and active line of work. At a high level, these results show that limiting the *information* revealed by earlier stages of analysis controls the *bias* introduced in later stages by adaptivity.

Here we review some known results in this area and highlight the role of information-theoretic concepts, notably several one-shot notions of mutual information.

1 Introduction

How can one do meaningful statistical inference and machine learning when data are *re-used* across analyses? The situation is common in empirical science, especially as data sets get bigger and more complex. For example, analysts often clean the data and perform various exploratory analyses—visualizations, computing descriptive statistics—before selecting how data will be treated. Many times the main analysis also proceeds in stages, with some sort of feature selection followed by inference using the selected features. In such settings, the analyses performed in later stages are chosen *adaptively* based on the results of earlier stages that used the same data. Adaptivity comes into even sharper relief when data are shared across multiple studies, and the choice of the research question in subsequent studies may depend on the outcomes of earlier ones. Adaptivity has been singled out as the cause of a "statistical crisis" in science [27].

There is a large body of work in statistics and machine learning on preventing false discovery, for example by accounting for multiple hypothesis testing. Classical theory, however, assumes that the analysis is fixed independently of the data—it breaks down completely when analyses are selected adaptively. Natural techniques, such as separating a validation set ("holdout") from the main data set to verify conclusions or the bootstrap method, do not circumvent the issue of adaptivity: once the holdout has been used, any further hypotheses tested using the same holdout will again depend on earlier results. Blum and Hardt [8] point out that this issue arises with leaderboards for machine learning competitions: they observe that one can do well on the leaderboard simply by using the feedback provided by the leaderboard itself on an adaptively selected sequence of submissions—that is, without even consulting the training data!

To formalize our situation somewhat, imagine there is a population that we wish to study, modeled by a probability distribution $\mathcal{P}$. An analyst selects a sequence of analyses $M_1, M_2, \ldots$, that she wishes to perform (we specify the type of analysis to consider later). In an ideal world (Figure 1, left), the analyst would run each analysis M_i on a fresh sample $\mathbf{X}^{(i)}$ from the population. For simplicity, we only discuss i.i.d samples in this article; we may assume that each sample $\mathbf{X}^{(i)}$ has n points, drawn independently from $\mathcal{P}$. In the real (adaptive) setting, the same data set $\mathbf{X}$ gets used for each analysis (Figure 1, right). The challenge is that we ultimately want to learn about $\mathcal{P}$, not $\mathbf{X}$, but adaptive queries can quickly overfit to $\mathbf{X}$.

Our goal is to relate these two settings—to develop techniques that allow us to emulate the ideal world in the real one, and understand how much accuracy is lost due to adaptivity. As mentioned above, merely setting aside a holdout to verify results at each stage is not sufficient, since the holdout ends up being re-used adaptively. If the number k of analyses is known ahead of time, one can split the data into k pieces of n/k points each (assuming i.i.d. data, the pieces are independent). This practice, called *data splitting*, provides clear validity guarantees, but is inefficient in its use of data: data splitting requires n to be substantially larger than k, while we will see techniques that do substantially better. Data splitting also requires an agreed upon partition of the data, which can be problematic with data shared across studies.

A line of work in computer science [21, 28, 20, 19, 40, 41, 6, 39, 46, 23, 24] initiated by Dwork, Feldman, Hardt, Pitassi, Reingold, and Roth [21] and Hardt and Ullman [28] provides a set of tools and specific methodology for this problem. This article briefly surveys the ideas in these works, with emphasis on the role of several information-theoretic concepts. Broadly, there is a strong connection between the extent to which an adaptive sequence of analyses remains faithful to the underlying population $\mathcal{P}$, and the amount of information that is leaked to the analyst about $\mathbf{X}$. In particular, randomization plays a key role in the state of the art methods, with a notion of algorithmic stability—*differential privacy*—playing a central role.

Another approach, with roots in the statistics community, seeks to model particular sequences of analyses, designing methodology to adjust for the bias due to conditioning on earlier results (e.g., [36, 30, 26, 22, 34, 32]). The specificity of this line of work makes it hard to compare with the more general approaches from computer science. Other work in statistics hews an intermediate path, allowing the analyst freedom within a prespecified class of analyses [7, 10]. There are intriguing similarities between these lines of work and the work surveyed here, such as the use of randomization to break up dependencies (e.g., [42, 43, 29]); understanding these connections more deeply is an important direction for future work.

2 The Lessons of Linear Queries

A simple but important setting for thinking about adaptivity, introduced by Dwork et al. [21] and Hardt and Ullman [28], is that of an analyst posing an adaptively selected sequence of queries, each of

*Computer Science and Engineering Department, Pennsylvania State University, University Park, PA, USA. asmith@psu.edu. Supported by NSF award IIS-1447700, a Google Faculty Award and a Sloan Foundation research award.

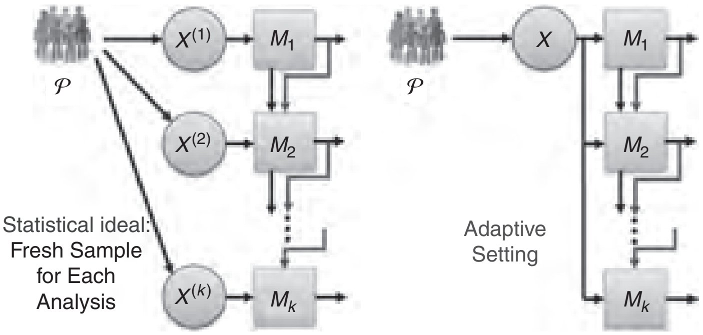


Figure 1. Ideally, we would collect fresh data from the same population $\mathcal{P}$ for each analysis (left). In many real settings, we have only a single data set that must be re-used, leading to adaptively selected analyses (right). The arrows pointing into the top of the analyses indicate that each analysis is selected based on the results of previous stages.

which asks for the expectation of a bounded function in the population. Such queries capture a wide range of basic descriptive statistics (the prevalence of a disease in a population, for example, or the average age). Many inference algorithms can also be expressed in terms of a sequence of such queries [31]; for example, optimization algorithms that query the gradient of a Lipschitz, decomposable loss function.

Suppose each data point lies in a universe $\mathcal{X}$, so that a data set lies in $\mathcal{X}^n$ and the underlying population is a distribution on $\mathcal{X}$. A *bounded linear query* is specified by a function $\phi : \chi \to [0,1]$. The *population value* of a linear query is simply the expected value of the function when evaluated on an element of the data universe drawn according to $\mathcal{P}$, denoted $\phi(\mathcal{P}) = \mathbb{E}_{x \sim \mathcal{P}}[\phi(x)]$.[1]

Consider now an interaction between an analyst wishing to pose such queries and an algorithm M (called the mechanism) holding a data set $\mathbf{X}$ sample i.i.d from $\mathcal{P}$ that attempts to provide approximate answers to the queries $\phi_1, \phi_2, \ldots$. This is illustrated in Figure 2, where we use subscripts (as in $M_1, M_2, \ldots$) to distinguish k different rounds of M. In general, neither the mechanism nor the analyst knows the exact distribution $\mathcal{P}$ (otherwise, why collect data?), so the mechanism cannot always answer $\phi(\mathcal{P})$. A natural approach is to answer with the *empirical mean* $\phi(X) = (1/n)\sum_{x_i \in x} \phi(X_i)$ When queries are selected nonadaptively, this is the best estimator of $\phi(P)$. We shall see, however, this is not the best mechanism for estimating the expectations of adaptively selected queries!

Given a query answering mechanism M, a data analyst A, and a distribution $\mathcal{P}$ on the data universe $\mathcal{X}$, consider a random interaction defined by selecting a sample $\mathbf{X}$ of n i.i.d. draws from $\mathcal{P}$, and then having A interact with $M(\mathbf{X})$ for k rounds, where in each round i, (i) A selects ϕ_i (based on $a_1, \ldots, a_{i-1}$), (ii) M answers a_i. The *(population) error* of M is the random variable

$$\mathrm{err}_X(M, A) = \max_i |\phi_i(\mathcal{P}) - a_i|.$$

which depends on $\mathbf{X}$ as well as the coins of M and A.

Definition 1. *A query answering mechanism $\mathcal{M}$ is (α,β)-accurate on i.i.d. data for k queries if for every data analyst A and distribution $\mathcal{P}$, we have*

$$\Pr(\mathrm{err}_x(M, A) \le \alpha) \ge 1 - \beta.$$

The probability is over the choice of the dataset $X \sim_{i.i.d.} \mathcal{P}$ and the randomness of the mechanism and the analyst. Similarly, the expected error *of M is the supremum, over distributions $\mathcal{P}$ and data analysts A, of $\mathbb{E}(\mathrm{err}_X(M, A))$. We sometimes fix the distribution $\mathcal{P}$ and take the supremum only over analysts A.*

It is important to note that this definition makes no assumptions on how the analyst selects queries, except that the selection is based on the outputs of M and not directly on the data. The aim of this line of work is to design mechanisms with provable bounds on accuracy. We aim for mechanisms that are *universal*, in the sense that they can be used in any type of exploratory or adaptive workflow.

2.1 Failures of Straightforward Approaches

As mentioned above, there are a couple of natural approaches to this problem. The first is to answer queries using each query's empirical mean $\phi(X)$. When queries are specified *non*adaptively, a standard argument shows that the population error of that strategy is $\Theta\left(\sqrt{(\log k)/n}\right)$

In contrast, in the adaptive setting, the empirical mechanism's error may be unbounded even with just two queries. For example, the query ϕ may be selected such that the low-order bits of $\phi_1(\mathbf{x})$ reveal all the entries of the data set $\mathbf{x}$. In that case, the analyst may construct a query ϕ_2 which takes the value 1 for values in the data set $\mathbf{x}$, and 0 otherwise. The empirical mean $\phi_2(\mathbf{x})$ will be 1, while the population mean $\phi_2(\mathcal{P})$ will be close to 0 for any distribution $\mathcal{P}$ with sufficiently high entropy.

This last example seems contrived, since it requires seemingly atypical structure from the initial query $\phi 1$. For example, constraining the queries ϕ to be *predicates* taking values in $\{0,1\}$ seems to eliminate the problem. However, the example is instructive for at least two reasons. First, it illustrates the role that *information about the data set* can play: learning $\mathbf{x}$ allows the analyst to pose a query that is highly overfit to the data set, and thus difficult for the mechanism to answer accurately. Conversely, we will see that limiting the information revealed about the data strongly limits overfitting.

Second, when the analyst asks more queries, one can construct much more natural examples of analyses that go awry when using the empirical mean. For instance, consider a data set where each k-individual data point lies in $\chi = \{0,1\}^{k-1} \times \{0, 1\}$, where we think of the first $k - 1$ bits as a vector of binary features, and the last bit as a label. Consider a particular analyst (from [20]) aiming to find a good classification rule for the label. The analyst's first $k - 1$ queries ask for the success rate of each of the $k - 1$ features in predicting the label. In the k-th query, the analyst constructs a classifier that takes a majority vote among those features that had success rate greater than 50%. On uniformly random data (where the label is independent of the

[1] The "linear" in "bounded linear query" refers to the fact that we care about the expectation of a function, so the resulting functional is a linear map from the set of distributions om X to [0, 1]. In contrast, some statistics, such as the variance of a random variable, are not linear. "Bounded" refers to the image being limited to [0, 1] (or, equivalently, any other finite interval).

features), the mechanism will report the success rate of this last classifier to be 55% when $k = (n/10)$, and 67% when $k = n$ (even though its success rate on the population would be 1/2). Generalizing the example somewhat, one can show that even with very simple data distributions, the error of empirical mechanism scales as $\Theta\left(\sqrt{k/n}\right)$—exponentially larger than the error one gets with nonadaptively specified queries. Encapsulating this discussion, we have:

Proposition 1. *When answering k nonadaptively specified queries, the empirical mechanism has expected error* $\Theta\left(\sqrt{\log(k)/n}\right)$ *When answering k adaptively selected queries, the empirical mechanism has expected error* $\Omega\left(\sqrt{k/n}\right)$, even for predicate queries on uniformly random data in $\{0, 1\}^k$.

Data Splitting Another natural approach for handling adaptively specified queries is *data splitting*: when k is known in advance, one may divide the data set into k subsamples of n/k points each, and answer the i-th query using its empirical mean on the i-th data set. This approach means that we can truly ignore adaptivity and use all the tools of classical statistics; the downside is that we are limited to the accuracy one can get with sample size n/k. The fact that we want a bound that is uniform over all k queries adds a further logarithmic factor to the final error bound:

Proposition 2. *When answering k adaptively specified queries, the data splitting mechanism has expected error* $\Theta\left(\sqrt{k(\log k)/n}\right)$.

For both of these natural mechanisms, answering queries with error α, even with constant probability, requires n to grow at least as fast as k/α^2. Can we do better? How good a dependency on α and k is possible?

2.2 A Sample of Known Bounds

In fact, there are mechanisms that can answer a sequence of k adaptively selected linear queries with much higher accuracy than that provided by the straightforward approaches. Namely, for a given accuracy α, we can get mechanisms that work for n that scales only as $\sqrt{k}/\alpha^2$—a quadratic improvement in k. The bounds below are stated in terms of expected error for simplicity; the underlying arguments also provide high-probability bounds on the tail of this error.

Theorem 3 ([21, 6])**.** *There is a computationally efficient mechanism for k statistical queries with expected error* $O\ (\sqrt[4]{k}\ /\sqrt{n})$.

A simple mechanism that achieves this bound is one that adds Gaussian noise with standard deviation about $\sqrt[4]{k}\ /\sqrt{n}$ to each query.

One can give a different-looking mechanism—which we do not describe in this survey—to automatically adjust to the actual "amount" of adaptivity in a given sequence of queries. Specifically, imagine that the k queries are grouped into r batches, where the queries in a given batch depend on answers to queries in previous batches but not on the answers to queries in the same batch. For example, in the classification example of the previous section, the number of rounds r is only 2.

Theorem 4 ([21])**.** *If there are at most r rounds of adaptivity, then there is a computationally efficient mechanism with expected error* $O\left(\sqrt{r(\log k)/n}\right)$ *The algorithm is* not *given the partition of the queries into batches.*

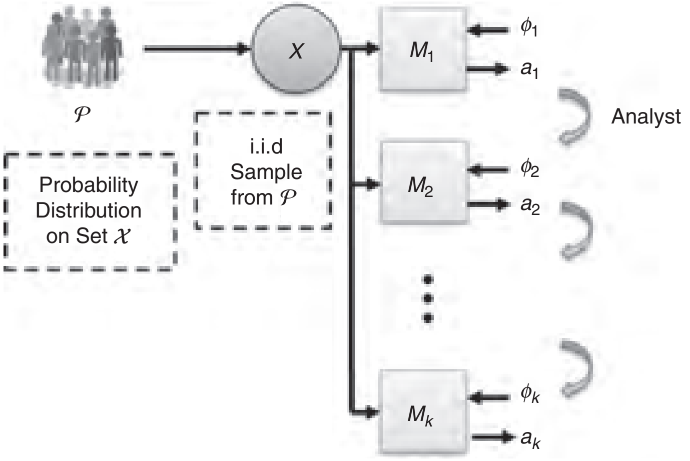


Figure 2. Adaptively selected linear queries.

The ideas underlying the two previous algorithms can also be adapted to give better results when we make further assumptions about the class of allowed queries, or the universe from which the data are drawn. One such result, due to Dwork et al. [21] (and tightened in [6]), recovers a logarithmic dependence on k in exchange for a dependence on the size of the universe χ in which the data lie.

Theorem 5 ([21, 6])**.** *There is a computationally inefficient mechanism with expected error* $O\left(\sqrt[6]{\log|\chi|}\sqrt[3]{(\log k)/n}\right)$. *The mechanism runs in time linear in* $|\chi|$ *(and not* $\log|\chi|$ *as one would naturally want).*

None of these upper bounds is known to be tight in all parameter regimes, but some lower bounds are known, in particular showing that the scaling $n = \Omega\ (\sqrt{k})$ cannot be improved, and that inefficiency of the mechanism in Theorem 5 is necessary.

Theorem 6 (Hardt and Ullman [28], Steinke and Ullman [41]). *For every mechanism M that answers k adaptively selected linear queries, for a sufficiently large universe* χ *(with log* $|\chi|$ *exponential in n), there exist a distribution* $\mathcal{P}$ *and an analyst* $\mathcal{A}$ *for which the mechanism's error* $\mathrm{err}_x(M, A)$ *is* $\Omega(\sqrt{k}/n)$ *with constant probability. Furthermore, for mechanisms that answer faithfully with respect to both the distribution and the data set (that is, they provide answers close to both* $\phi_i(\mathbf{X})$ *and* $\phi_i(\mathcal{P})$*), the bound can be strengthened to* $\Omega(\sqrt[4]{k}/\sqrt{n})$

Finally, if we assume that one-way functions exist, then the bounds continue to hold when log $|\chi|$ *has polynomial size, for polynomial-time mechanisms (but not for those that can take exponential time).*

We won't discuss the proof of these lower bounds here, but we note that closing the gap between the upper and lower bounds remains an intriguing open problem.

2.3 Privacy and Distributional Stability

The upper bounds above are obtained via a connection between adaptive analysis and certain notions of algorithmic stability. Broadly, algorithmic stability properties limit how much the output of an algorithm can change when one of its inputs is changed. Different notions of stability correspond, roughly, to different measures of distance between outputs. There is a long-standing connection between algorithmic stability and expected generalization error (e.g., Devroye and Wagner [14], Bousquet and Elisseeff [9]). Essentially, *stable algorithms cannot overfit.* It seems that if we could design adaptive query-answering mechanisms that are stable in an appropriate sense, we could get validity guarantees for adaptive data analysis.

Alas, there is a hitch. Recall that our goal is to design mechanisms that provide statistically valid answers no matter how the analyst selects queries. Even if each stage of the mechanism is stable, the overall process might not be—in an adaptive setting, the analyst ends up being part of the mechanism.

The resolution is to consider a *distributional* notion of stability. We will require that changing any single data point in x have a small effect on the distribution of the mechanism's outputs. If we choose a distance measure on distributions that is nonincreasing under postprocessing, then we can limit the effect of the analyst's choices.

Specifically, we work with "differential privacy," a notion of stability introduced in the context of privacy of statistical data. Differential privacy seeks to limit the information revealed about any single individual in the data set.

Definition 2 ([17, 16]). *An algorithm* $M:\chi^n \to \mathcal{O}$ *is* (ϵ, δ)*-differentially private if for all pairs of neighboring data sets* $\mathrm{x}, \mathrm{x}' \in \chi^n$*, and for all events* $S \subseteq \mathcal{O}$*:*

$$\Pr[M(\mathrm{x}) \in S] \leq \exp(\epsilon)\Pr[M(x') \in S] + \delta.$$

Differential privacy makes sense even for interactive mechanisms that involve communication with an outside party: we simply think of the outside party as part of the mechanism, and define the final output of the mechanism to be the complete transcript of the communication between the mechanism and the other party.

Differential privacy is a useful design tool in the context of adaptive data analysis because it is possible to design interactive differentially private algorithms *modularly,* due to two related properties: closure under postprocessing, and composition:

Proposition 7. *If* $M:\chi^n \to \mathcal{O}$ *is* (ϵ, δ)*-differentially private, and* $f:\mathcal{O} \to \mathcal{O}'$ *is an arbitrary (possibly randomized) mapping, then* $f \circ M:\chi^n \to \mathcal{O}'$ *is* (ϵ, δ)*-differentially private.*

Proposition 8 (Adaptive Composition [18, 35]—informal). *Let* M_1*,* M_2*, ...,* M_k *be a sequence of* (ϵ, δ)*-differentially private algorithms that are all run on the same data set, and selected adaptively (with the choice of* M_i *depending on the outputs of* M_1*, ...,* M_{i-1}*, but not directly on* **x**). *Then no matter how the adaptive selection is done, the resulting composed process is* (ϵ', δ')*-differentially private, for* $\epsilon' \approx \epsilon\sqrt{k}$ *and* $\delta' \approx k\delta$.

Taken together, these two properties mean that in order to design differentially private algorithms for answering linear queries, it is sufficient to make sure the mechanism run at each stage is differentially private.

Perhaps even more importantly, in order to ensure statistical validity—that is, accuracy with respect to the underlying population—it suffices to design differentially private algorithms that are accurate with respect to the sample x:

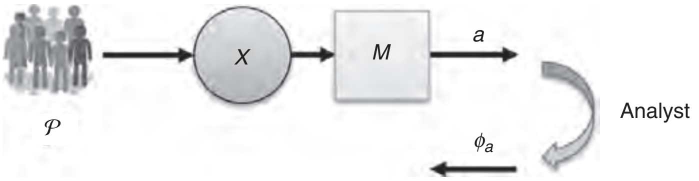


Figure 3. A two-stage overfitting game.

Theorem 9 (Main Transfer Theorem [21, 6]). *Suppose a statistical estimator M is* $(\varepsilon, \varepsilon \cdot \delta)$*-accurate with respect to its sample, that is, for all data sets* x,

$$\Pr\left(\max_i |a_i - \phi_i(\mathrm{x})| \leq \epsilon\right) \geq 1 - \epsilon \cdot \delta.$$

If M is also $(\epsilon, \epsilon \cdot \delta)$*-differentially private, then it is* $(O(\epsilon), O(\delta))$*-accurate with respect to the population* (*Definition* 1).

This theorem underlies two of the three upper bounds of the previous section (Theorems 3 and 5). Each is derived by using existing differentially private algorithms together with Theorem 9. For Theorem 4, Dwork et. al. [21] used a different argument, based on compressing the output of the algorithm to a small set of possibilities; see Section 3.

2.4 A Two-stage Game, Stability and "Lifting"

We conclude this section with an outline of the proof of the main transfer theorem (Theorem 9). That theorem talks about, analyses with many stages of interaction, but it turns out that the core of the argument lies in understanding a seemingly much simpler, two-stage process.

Consider a two-stage setting in which an analysis M is run on data set x, and the analyst, selects a linear query $\phi : \chi \to [0, 1]$ based on $M(\mathbf{x})$ (Figure 3). We say *M robustly* (α, β)*-generalizes* if for all distributions $\mathcal{P}$ over the domain χ, for all strategies (functions) A employed by the analyst, with probability at least $1 - \beta$ over the choice of $\mathbf{X} \sim \mathcal{P}^n$ and the coins of M, we have that $|\phi(\mathrm{X}) - \phi(\mathcal{P})| \leq \alpha$ where $\phi = A(M(\mathbf{X}))$. (Similarly, we may talk about, the expected generalization error, that is, the maximum over A and $\mathcal{P}$ of $\mathbb{E}(|\phi(\mathrm{X}) - \phi(\mathcal{P})|)$.)

The quantification over all selection functions A here is critical—when the first, phase of analysis satisfies the definition, then a query asked in the following round cannot, overfit to the data, (except, with low probability), no matter how it is selected.

Differential privacy (and a few other distributional notions of stability, such as KL-stability [6, 47]) limits the adversary's score in this game. This connection had been understood for some time— for example, McSherry observed that it could be used to break up dependencies in a clustering algorithm, and Bassily, Smith, and Thakurta [5] used a weak version of the connection to bound the population risk of differentially private empirical risk minimization.

However, the application to adaptive data analysis-and especially the understanding of the importance of post-processing to the design of universal mechanisms—came recently in [21]. Their initial result was subsequently sharpened, to obtain the following tight connection:

Theorem 10 (Differentially Private Algorithms Cannot Overfit [6]). *If M is $(\epsilon, \epsilon\delta)$-differentially private, then it is $(O(\epsilon), O(\delta))$-robustly generalizing.*

Lifting to Many Stages Bounds on the two-stage game can be "lifted" to provide bounds on the *k*-phase game either through a sequential application of the bound to each round [21] or through a more holistic argument, called the *monitor technique* [6], that yields Theorem 9 (and Theorems 3 and 5).

The monitor argument is a thought experiment—we argue that for any multi-stage process, there is a two-stage process in which the error on the population equals the maximum population error over all *k* stages of the original process. The argument applies quite generally, but it is a bit simpler to explain under the assumption that the mechanism answers queries accurately with respect to the data set **X**. The idea, given an interaction between an analyst and adaptively selected mechanisms $M_1, M_2 \ldots, M_k$, is to encapsulate the analyst and mechanisms into a single fictional entity M which gets, as additional input, the underlying distribution $\mathcal{P}$. The fictional M executes an interaction and then outputs a single query ϕ^*—the one which maximizes the population error over all stages i.

Beyond Linear Queries The techniques described in this section extend to problems that are not described by estimating the mean of a bounded linear functional. One important class is minimizing a decomposable loss function where each individual contributes a bounded term to the loss function [6].

A Re-usable Holdout The techniques described in this section can appear somewhat onerous for the analyst, since they require accessing data via differentially private algorithms. As pointed out by Dwork et al. [20], however, one need not limit, access to the entire data set in this way. In fact, a more pragmatic approach is to give most, of the data, "in the clear" to the analyst, and protect, only a small holdout set via the techniques discussed here This still allows one to *verify* conclusions soundly, but additionally allows full exploratory analysis, as well as repeated verification ("holdout re-use").

3 The Intrigue of Information Measures

Despite the generality of the approach of the previous section, many important classes of analyses are not obviously amenable to those techniques; in particular, problems that are not easily stated in terms of a numerical estimation task.

Consider the problem of hypothesis testing. Crudely, given a set of distributions $\mathcal{H}$ (called the null hypothesis), we ask if the data set is "unlikely to have been generated" by a distribution $P \in \mathcal{H}$. More precisely, we select an event T (the *acceptance region*) such that $\Pr_{X \sim \mathcal{P}^n}(X \in T)$ is at most a threshold γ (often 0.05) for all distributions in $\mathcal{H}$. If it happens that the observed data **X** lie outside of T, the null hypothesis is said to be rejected. If this happens when the true distribution $\mathcal{P}$ is actually in $\mathcal{H}$, then we say a *false discovery* occurs. Hypothesis tests play a central role in modern empirical science (for better or for worse), and techniques to control false discovery in the classic, nonadaptive setting are the focus of intense study. Despite this, very little is known about hypothesis tests in adaptive settings.

Adaptivity arises when the event T is selected based on earlier analysis of the same data—conditioned on those earlier results $Y = M(\mathbf{X})$, the probability that the test rejects the null hypothesis given **X** might be much higher than γ even if $\mathcal{P}$ lies in $\mathcal{H}$.

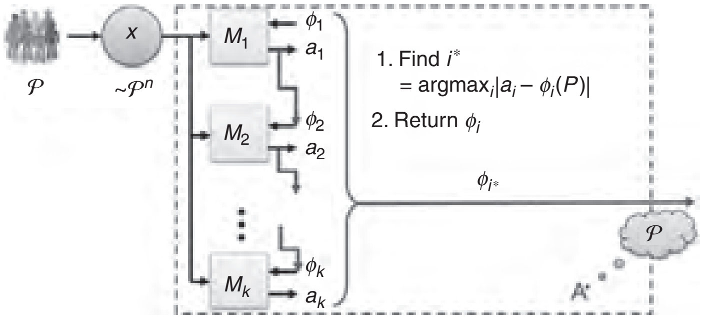


Figure 4. The "monitor" argument for lifting results about a two-stage game to many stages.

How much higher it can be depends on M and—as we will see—on several measures of the information leaked by M. To formalize this, consider a game similar to the overfitting game, in which the analyst A, given $Y = M(\mathbf{X})$, selects an arbitrary event $T_Y = A(Y)$ (which depends on Y). For a particular output Y of M, the analyst's "score" is

$$\text{score}_{\mathcal{P},M,A,\gamma}(y) = \begin{cases} \Pr_{\mathbf{X}\sim\mathcal{P}^n}(\mathbf{X} \in T_y | Y = y) & \text{if} \quad \Pr_{\mathbf{X}\sim\mathcal{P}^n}(\mathbf{X} \in T_y) \le \gamma, \\ 0 & \text{if} \quad \Pr_{\mathbf{X}\sim\mathcal{P}^n}(\mathbf{X} \in T_y) > \gamma. \end{cases}$$

Now consider the analyst's *expected* score in this game: $\eta_{\mathcal{P},M,A}(\gamma) = \mathop{\mathbb{E}}_{Y}(\text{score}_{\mathcal{P},M,A}(y))$. As we will see below, the analyst's score in this game can be bounded using various definitions of the *information* leaked about **X** by Y. This score also plays a key role in controlling false discovery:

Proposition 11. *Bounding the score η has several important implications:*

1) (False discovery [39])[2] *If $Y = M(\mathrm{X})$ is used by A to select a hypothesis test with significance γ, then the probability of false discovery is at most $\eta_{\mathcal{P},M,A}$.*

2) (Robust generalization [19]) If $Y = M(X)$ is used by A to select a bounded linear query ϕ_Y, then $\Pr(|\phi_Y(\mathrm{X}) - \phi_Y(\mathcal{P})| \ge t/\sqrt{n}) \le \eta_{\mathcal{P},M,A}(e^{-t^2/3})$

We are interested in universal bounds that hold no matter how the analyst uses the output Y, and no matter the original input distribution. To this end, we define

$$\eta_M(\gamma) = \sup_{\mathcal{P},A}(\eta_{\mathcal{P},M,A}(\gamma))$$

3.1 *Information and Conditioning*

The function η_M measures how much probabilities less than γ can be amplified by conditioning on Y, on average over values of Y.

For several one-shot notions of mutual information, we have that if a procedure leaks k "bits" of information we have $\eta_{\mathcal{P},M,A}(\gamma) \approx \gamma \cdot 2^k$.

[2] The definitions here differ somewhat from those of [39]; in particular, our function η is the inverse of a "p-valuecorrection function" from [39].

Unfortunately, such a clean relationship is not known for the standard notion of Shannon mutual information. Instead, we consider two other notions here.

Fix two random variables X, Y with joint distribution given by p_{XY} $(x, y) = \Pr(X = x, Y = y)$ and marginals $p_X(\cdot)$ and $p_Y(\cdot)$. Consider the information loss

$$I_{x,y} = \log\left(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right).$$

The standard notion of mutual information is the expectation of this variable: $I(X;Y) = \mathbb{E}\,(I_{X,Y})$. The *max-information* [12, 19] between X and Y is the supremum of this variable. Unfortunately, for many interesting procedures, the max-information is either unbounded or much larger than the mutual information.

One can get a more flexible notion by instead considering a high-probability bound on the information loss: we say the *β-approximate max information* between X and Y is at most k (written $I_\infty^\beta(X{:}Y) \le k$) if $\Pr_{(x,y)\sim(x,y)}(I_{x,y} \le k) \ge 1-\beta$.[3,4]

For many algorithms of interest, the approximate max-information turns out to be very close to the mutual information but, by providing a bound on the upper tail of *Ix,y*, allows for more precise control of small-probability events.

We can also define a related quantity, which we call the *expected log-distortion:*

$$L_\infty(X;Y) = \log \mathop{\mathbb{E}}_{y\sim Y}\left(\sup_x (2^{I_{x,y}})\right) = \log \mathop{\mathbb{E}}_{y\sim Y}\left(\sup_x \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\right).$$

This notion of information leakage is not symmetric in X, Y. It is closely related to, but in general different from, the min-entropy leakage $H_\infty(Y|X) - H_\infty(X)$ [15, 1, 25, 2, 3].

Of these notions, expected log distortion is the strongest since it upper bounds the other two: $I(X;\ Y) \le L_\infty(X;\ Y)$ and $I_\infty^\beta(X;Y) \le L_\infty(X;Y) + \log(1/\beta)$.

Theorem 12 *For every mechanism M and distribution $\mathcal{P}$ s.t.* $\mathbf{X} \sim \mathcal{P}$:

1) If $I_\infty^\beta(X;Y) \le k$, *then for every analyst A,* $\eta_{\mathcal{P},A,M}(\gamma) \le \gamma \cdot 2^k + \beta$, *and M robustly* $(\alpha,2\beta)$*-generalizes for* $\alpha = \sqrt{(k+2\ln(1/\beta))/n}$ *on* $\mathcal{P}$.

2) If $L_\infty(\mathbf{X}; M(\mathbf{X})) \le k$, *then for every analyst A,* $\eta_{\mathcal{P},A,M}(\gamma) \le \gamma \cdot 2^k$, *and for every $\beta > 0$, M robustly (α,β)-generalizes for* $\alpha = \sqrt{(k+2\ln(1/\beta))/n}$ *on $\mathcal{P}$.*

We know much weaker implications based only on bounding the mutual information. Most significantly, the bounds for general hypothesis testing are exponentially weaker than those one gets from the one-shot measures above

Proposition 13 ([40, 39]). If $I\,(X; M\,(X)) < k$, then (*i*) for every analyst A, $\eta_{p,M,A}(\gamma) \le (k + 1/\log_2(1/\gamma))$, and (ii) for every $\beta > 0$, M robustly (a, β)-generalizes for $\alpha = O\left(\sqrt{k/n\beta}\right)$ on $\mathcal{P}$.

3.2 What procedures have bounded one-shot information measures?

The information-theoretic framework of the previous subsection captures several other classes of algorithms that satisfy robust generalization guarantees. In addition to unifying the previous work, this approach shows that these classes of algorithms allow for principled post-selection hypothesis testing.

The most important of these, currently, is for the class of differentially private algorithms:

Theorem 14 (Informal, see [39]). *If M is (ϵ,δ)-differentially private, and the entries of* $\mathbf{X}$ *are independent, then* $I_\infty^\beta(\mathbf{X};M(\mathbf{X})) = O(\epsilon^2 n)$ *for* $\beta = O(n\sqrt{\delta/\epsilon})$.

This result, together with Theorem 12, implies that differentially-private algorithms are $(O(\epsilon), O(n\sqrt{\delta/\epsilon})$-robustly generalizing for data drawn i.i.d from any distribution $\mathcal{P}$. It essentially recovers the results of the previous section on linear queries (with a worse value of β), but additionally applies to more general problems such as hypothesis testing.

Description length [19] In many cases, the outcome of a statistical analysis can be compressed to relatively few bits—for example, when the outcome is a small set of selected features. If the output of M can be compressed to k bits, then the expected log-distortion L_∞ (X; M(X)) is at most k bits. An argument along these lines was used implicitly in [21] to prove Theorem 4.

Compression Schemes Another important class of statistical analyses that have good (and robust) generalization properties are compression learners [33]. These process a data set of n points to obtain a carefully selected subset of only $k << n$ points, and finally produce an output fit to those points. A classic example is support vector machines: in d dimensions, the final classifier is determined by just d + 1 points in the data set.

Cummings, Ligett, Nissim, Roth, and Wu [13] used classic generalization results for such learners to show that they satisfy robust generalization guarantees. The classic results as well as those of Cummings et al. [13] can be rederived from the following lemma (new, as far as we know) bounding the information leaked by a compression scheme about those points that are *not* output by the scheme.

Lemma 15 (Compression schemes). *Let* $M : \mathcal{X}^n \to \mathcal{X}^k$ *be any algorithm that takes a data set X of n points and outputs a subset* $\mathrm{x}_{\mathrm{out}} = M$ (x) *of k points from* x. *Let* $\mathrm{x}_{\mathrm{in}}\ 2 \in \mathcal{X}^{n-k}$ *denote the remaining data points, so that* $\mathrm{x}_{\mathrm{in}} \cup \mathrm{x}_{\mathrm{out}} = \mathrm{x}$ *(as multisets). For any distribution $\mathcal{P}$ on $\mathcal{X}$, if* X $\sim \mathcal{P}^n$*, then*

$$L_\infty(X_{in};X_{\mathrm{out}}) \le \log_2\binom{n}{k}.$$

Cummings et al. [13] used the robust generalization properties of compression learners to give robustly generalizing algorithms for learning any PAC-learnable concept class. In particular, this

[3] This condition is not exactly the definition of max-information of [19] (which requires that for all events E, $\Pr((X,Y) \in E) \le 2^k \Pr((X', Y) \in E)$ where X' is identically distributed to X but independent from Y). The definition here implies that of [21].

[4] The β-approximate max information is equivalent to a *smoothed* version of max-information [37, 38, 44, 45, 12], in which we ask that the pair X, Y be within statistical distance β of a joint distribution with $I_\infty^\beta(X;Y) \le k$. See Corollary 8.7 in Bun and Steinke [11] for details.

implies robustly generalizing algorithms for tasks that do not have differentially private algorithms, such as learning a threshold classifier with data from the real line.

Notes and Acknowledgments

I am grateful to many colleagues for introducing me to this topic and insightful discussions. Among others, my thanks go to Raef Bassily, Cynthia Dwork, Moritz Hardt, Kobbi Nissim, Ryan Rogers, Aaron Roth, Thomas Steinke, Uri Stemmer, Om Thakkar, Jon Ullman, Yu-Xiang Wang. Paul Medvedev provided helpful comments on the writing, making the survey a tiny bit more accessible to nonexperts. Finally, I thank Michael Langberg, this newsletter's editor, for soliciting this article and graciously tolerating delays in its preparation.

Although I tried to cover the main ideas in a recent line of work, that line is now diverse enough that I could not encapsulate everything here. Notable omissions include the "jointly Gaussian" models of Russo and Zou [40] and Wang et al. [46], which assume that the analyst is selecting among a family of statistics that are jointly normally distributed under $\mathcal{P}$, work of Elder [23, 24] on a Bayesian framework that encodes a further restriction that the mechanism "know as much" as the analyst about the underlying distribution $\mathcal{P}$, and the "typical stability" framework [4, 13].There are no doubt other contributions that I lost in my effort to consolidate. Some ideas in this survey, notably the information-theoretic viewpoint on compression-based learning, have not appeared elsewhere; they have not been peer-reviewed.

References

[1] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith. Measuring information leakage using generalized gain functions. In *25th IEEE Computer Security Foundations Symposium (CSF)*, pages 265–279, 2012.

[2] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith. Additive and multiplicative notions of leakage, and their capacities. In *27th IEEE Computer Security Foundations Symposium (CSF)*, pages 308–322, 2014.

[3] M. S. Alvim, K. Chatzikokolakis, A. McIver, C. Morgan, C. Palamidessi, and G. Smith. Axioms for information leakage. In *29th IEEE Computer Security Foundations Symposium (CSF)*, pages 77–92, 2016.

[4] R. Bassily and Y. Freund. Typical stability. arXiv:1604.03336 [cs. LG], April 2016.

[5] R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *IEEE Symposium on the Foundations of Computer Science (FOCS)*, pages 464–473, 2014.

[6] R. Bassily, K. Nissim, A. Smith, T. Steinke, U. Stemmer, and J. Ullman. Algorithmic stability for adaptive data analysis. In *48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1046–1059, 2016.

[7] R. Berk, L. Brown, A. Buja, K. Zhang, and L. Zhao. Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837, 2013.

[8] A. Blum and M. Hardt. The ladder: A reliable leaderboard for machine learning competitions. In *Proc.32 nd International Conference on Machine Learning*, 2015. arXiv:1502.04585.

[9] O. Bousquet and A. Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2: 499–526, 2002.

[10] A. Buja, R. Berk, L. Brown, E. George, E. Pitkin, M. Traskin, L. Zhao, and K. Zhang. Models as approximations—a conspiracy of random regressors and model deviations against classical inference in regression. *Statistical Science*, 1460, 2015.

[11] M. Bun and T. Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference (TCC) 2016-B*, 2016. arxiv:1605.02065.

[12] N. Ciganovic, N. J. Beaudry, and R. Renner. Smooth max-information as one-shot generalization for mutual information. *IEEE Trans. Information Theory*, 60(3):1573–1581, 2014.

[13] R. Cummings, K. Ligett, K. Nissim, A. Roth, and Z. S. Wu. Adaptive learning with robust generalization guarantees. In *29th Annual Conference on Learning Theory*, pages 772–814, 2016.

[14] L. Devroye and T. Wagner. Distribution-free performance bounds for potential function rules. *IEEE Transactions on Information Theory*, 25(5):601–604, 1979.

[15] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008.

[16] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT*, pages 486–503, St. Petersburg, Russia, 2006.

[17] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

[18] C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 51–60, Washington, DC, USA, 2010. IEEE Computer Society.

[19] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth. Generalization in adaptive data analysis and holdout reuse. In *Advances in Neural Information Processing Systems*, pages 2350–2358, 2015.

[20] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth. The reusable holdout: Pre- serving validity in adaptive data analysis. *Science*, 349(6248):636–638, 2015.

[21] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. L. Roth. Preserving statistical validity in adaptive data analysis. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 117–126. ACM, 2015.

[22] B. Efron. Estimation and accuracy after model selection. *Journal of the American Statistical Association*, 109(507):991–1007, 2014.

[23] S. Elder. Challenges in bayesian adaptive data analysis. arXiv:1604.02492, 2016.

[24] S. Elder. Bayesian adaptive data analysis guarantees from subgaussianity. arXiv:1611.00065 [cs.LG], 2016.

[25] B. Espinoza and G. Smith. Min-entropy as a resource. *Inf. Comput.*, 226:57–75, 2013.

[26] W. Fithian, D. Sun, and J. Taylor. Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*, 2014.

[27] A. Gelman and E. Loken. The statistical crisis in science. *American Scientist*, 102(6):460, 2014.

[28] M. Hardt and J. Ullman. Preventing false discovery in interactive data analysis is hard. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 454–463. IEEE, 2014.

[29] X. T. Harris, S. Panigrahi, J. Markovic, N. Bi, and J. Taylor. Selective sampling after solving a convex problem. *arXiv: 1609.05609*, 2016.

[30] C. M. Hurvich and C.-L. Tsai. The impact of model selection on inference in linear regression. *The American Statistician*, 44(3):214–217, 1990.

[31] M. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45 (6):983–1006, 1998.

[32] J. D. Lee, D. L. Sun, Y. Sun, and J. E. Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927, 2016.

[33] N. Littlestone and M. Warmuth. Relating data compression and learnability. Technical report, 1986.

[34] R. Lockhart, J. Taylor, R. J. Tibshirani, and R. Tibshirani. A significance test for the lasso. *The Annals of Statistics*, 42(2):413, 2014.

[35] S. Oh and P. Viswanath. The composition theorem for differential privacy. *CoRR*, abs/1311.0776, 2013. URL http://arxiv.org/abs/1311.0776.

[36] B. M. Pötscher. Effects of model selection on inference. *Econometric Theory*, 7(2):163–185, 1991.

[37] R. Renner and S. Wolf. Smooth Rényi entropy and applications. In *IEEE International Symposium on Information Theory — ISIT*, page 233, 2004.

[38] R. Renner and S. Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in Cryptology - ASIACRYPT*, pages 199–216, 2005.

[39] R. Rogers, A. Roth, A. Smith, and O. Thakkar. Max-information, differential privacy, and post-selection hypothesis testing. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, 2016.

[40] D. Russo and J. Zou. Controlling bias in adaptive data analysis using information theory. In *19th International Conference on Artificial Intelligence and Statistics*, pages 1232–1240, 2016. arXiv:1511.05219.

[41] T. Steinke and J. Ullman. Interactive fingerprinting codes and the hardness of preventing false discovery. In *Proceedings of The 28th Conference on Learning Theory*, pages 1588–1628, 2015.

[42] X. Tian and J. E. Taylor. Selective inference with a randomized response. *arXiv:1507.06739*, 2015.

[43] X. Tian, N. Bi, and J. Taylor. Magic: a general, powerful and tractable method for selective inference. *arXiv: 1607.02630*, 2016.

[44] M. Tomamichel, R. Colbeck, and R. Renner. Duality between smooth min- and max-entropies. *IEEE Trans. Information Theory*, 56(9):4674–4681, 2010.

[45] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner. Chain rules for smooth min- and max-entropies. *IEEE Trans. Information Theory*, 59(5):2603–2612, 2013.

[46] Y.-X. Wang, J. Lei, and S. E. Fienberg. A minimax theory for adaptive data analysis. arXiv:1602.04287 [stat.ML], 2016.

[47] Y.-X. Wang, J. Lei, and S. E. Fienberg. On-average KL-privacy and its equivalence to generalization for max-entropy mechanisms. arXiv:1605.02277 [stat.ML], 2016.

From the President

continued from page 1

This is not the only book project in the works by the way. I hope to have more to report in the next column.

By the time you read this, it will be time for ISIT in Aachen, Germany. No doubt it will be perfectly organized and smoothly run like a well-oiled engine, emitting a maximal amount of excitement. If you have ever been involved in organizing one of these large conferences, you know the incredible amount of effort that goes into such an event. Therefore, a great thank you to the organizing team! Hopefully their heroic efforts inspire you to check out the article written by Emanuele Viterbo, the head of our conference committee, and Elza Erkip entitled “Thinking about organizing an ITW or ISIT?”.

I hope to see many of you in Aachen!
Ruediger

IEEE Communications Society and Information Theory Society Joint Paper Award for 2017

The paper "List Decoding of Polar Codes", *IEEE Transactions on Information Theory*, Vol. 61, No. 5, pp 2213–2226, May 2015 by **Ido Tal and Alexander Vardy** has been awarded the 2017 IEEE Communications Society and Information Theory Society Joint Paper Award for 2017.

Congratulations!

IEEE Information Theory Society's New Distinguished Lecturers

The Information Theory Society established the Distinguished Lecturer Program to promote interest in information theory by supporting chapters who wish to invite prominent information theory researchers to give talks at their events. Distinguished Lecturers are selected by the Membership and Chapters (MC) Committee in consultation with the Board of Governors. The Society aims to maintain ten Distinguished Lecturers each serving for two year terms.

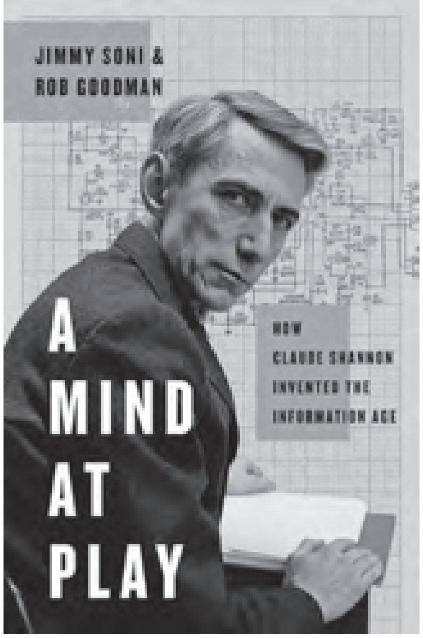
Congratulations to five new lecturers that have been named by the Society as Distinguished Lecturers for 2017–2018:

Tara Javidi, Navin Kashyap, Chandra Nair, Osvaldo Simeone and Ram Zamir.

Behind the Scenes: A Q&A with the Authors of Claude Shannon's New Biography

Jimmy Soni and Rob Goodman

We've spent the prior five years writing a biography about Claude Shannon. We are unexpected chroniclers of his life: We aren't engineers or mathematicians or even scientists; we are biographers (our first book was about an ancient Roman senator named Cato). It took a good deal of research—and lots of help from people more knowledgeable than us!—to put together what we think is a compelling window into the life of one of the 20th century's great minds.



Many people who work in information theory have, understandably, wondered how two non-technical writers came to this project. We figured we would share how we got here, what we learned along the way, and how this book came to be:

How Did You Come to Learn About Claude Shannon? What Got You Hooked?

The idea came from a friend, who sent Jimmy a copy of Jon Gertner's book, *The Idea Factory,* a narrative history of the Bell Laboratories. The gift included a note that said, half-jokingly, "Your next book ought to be about Claude Shannon." Jimmy found Shannon to be an engaging figure. He went looking for a biography of him and couldn't find one. And thus a book idea was born!

Our agent connected us with Alice Mayhew at Simon & Schuster, the editorial force behind *Steve Jobs* and *Einstein* by Walter Isaacson and *A Beautiful Mind* by Sylvia Nasar. She's shepherded books of this kind before, and she understood this idea instantly. Five years later, we have ourselves a book. And as interesting as we thought Shannon was when we knew only the barest details of his life and work, we find him all the more compelling now. He was a once-in-a-generation figure, and we hope our biography manages to capture that.

What Did You Learn About Him That Surprised You?

Those who know Shannon's work appreciate its rigor, thoughtfulness, and depth. Those are his hallmarks and they are why he still commands such respect today. What's less understood, or what surprised us, at least, was Shannon's artistic and creative bent. He had a flair for the visual—and it's telling that, alongside the landmark papers published in journals, he constructed objects that are, to this day, featured in museums. His son, spoke with us at length about this, probably because he was one of the few who had a front-row seat to all the tinkering and building that Shannon did as an adult.

Many of the objects Shannon made have a theatrical quality: The fire-breathing trumpet, for instance, or the Ultimate Machine, whose only purpose is to turn itself off. Even Theseus the maze-solving mouse, his best-known creation, is designed for maximum visual impact: the mouse is just an accessory to the real "artificially intelligent" system, the 75 relays that operate below the maze and allow the mouse to locate a metal piece of cheese. His contraptions may have been private curiosities, but they long for an audience—and there's real artistry and creativity baked into this work. That same artistry, we would contend, drove Shannon's most groundbreaking theoretical work, including his innovations in information theory. An engineer we spoke with called this quality "thinking not only *about* things, but *through* things."

Who Do You Think Influenced His Work?

Shannon was the sort of student whose talent marked him out almost immediately as a promising protege, and he crossed paths with many of the most influential scientists and engineers of his day. Perhaps his most important early mentor was Vannevar Bush, the MIT engineering professor who went on to play a crucial role in coordinating America's scientists during WWII. It was Bush who hired Shannon to work on MIT's differential analyzer, a massive analogue computer, as Shannon completed his Master's degree; and it was during his work for Bush that Shannon wrote his famous thesis on the use of digital switches for Boolean logic, which Walter Isaacson called "the basic concept underlying all digital computers." Another important early influence for Shannon was his dissertation supervisor, the geneticist Barbara Stoddard Burks, who oversaw his work developing what he called "an algebra for theoretical genetics."

Shannon also drew inspiration from the histories of the fields he worked in. "A Mathematical Theory of Communication," his landmark work on information theory, cites the earlier work of Harry Nyquist and Ralph Hartley in developing the scientific concept of information; our book describes how their work at Bell Labs shaped the field as Shannon found it. In another unpublished paper on, of all things, juggling, he cites a wide range of sources: the science fiction author Robert Silverberg, Captain James Cook, Xenophon, W.C. Fields, and the jazz drummer Gene Krupa, among many others—which should give you an idea of just how eclectic Shannon's influences and interests were. Shannon read very widely, and he found inspiration in equal parts from T.S. Eliot (his favorite poet) and Lewis Carroll. He was moved by music and art as much as by mathematics and engineering, and it's important to include those among his influences.

What was the Process of Writing the Biography?

It involved a mix of in-depth research and interviews with Shannon's family, friends, and associates. We were fortunate to have strong guides through the world of information theory, people like Dr. Sergio Verdú, as well as access to the people who knew Shannon and his work. Andrew and Peggy Shannon, his son and daughter, were very generous with their time. Betty Shannon opened up to us about the Claude she knew. Robert Gallager, Len Kleinrock, Henry Pollak, Bob Fano, Ed Thorp, and many others gave us vivid descriptions of what it was like to study under or work with Shannon. The interviews were the most useful, and the most enjoyable, part of writing the book. It was a pleasure to absorb Shannon lore from the people who experienced it first-hand, and to help preserve it for posterity.

Having Closely Studied His Life, What Do You Think is the Key to His Breakthroughs?

It's hard to point to one thing. There are a few elements that run through his body of work, and each one, we think, is important.

He was a tinkerer from his earliest days. As a boy, he built a barbed-wire telephone line and a makeshift elevator in a friend's barn; his hands were always busy, taking things apart and putting them back together. He brought this tinkering spirit to much of his work, building with his hands to test and refine what he dreamt up with his mind.

Shannon was persistent, especially early in his career. Even though he gave the impression of a carefree scholar, Shannon was phenomenally hard-working, and he stuck with problems long past the point at which others might have given up. Shannon's work on information theory is so elegant that we can sometimes overlook the fact that his "Mathematical Theory of Communication" was developed over the course of a decade—and he did the bulk of the work on it in his spare time.

There was, as our title suggests, an element of play throughout the work. You get the impression that each of these intellectual puzzles were a joy to figure out rather than a bore. Codebreaking, chess-playing machines, information theory—it all came from the same joyful place.

And finally, he had some gifts and some luck. A gift for simplification, for instance, which led him to the essence of things. He'd also been fortunate: fortunate to have earned the early backing of Vannevar Bush; lucky to have found his way to Bell Labs, which might have been the one place on earth that would tolerate him; blessed to have spent World War II on cryptography and the mathematics of anti-aircraft fire rather than fighting in combat.

Why Do You Think He Doesn't Have the Name Recognition of an Einstein or a Newton?

Partly because he didn't chase the attention. Shannon had numerous opportunities to become a scientific celebrity. In fact, in the 1950s and 1960s, he got a taste of it: he made the rounds of media and was written up in *Life* and *Time Magazine*. But playing the part of celebrity intellectual was an odd fit for him, and he gave it up almost as soon as he got it. He preferred to follow his own curiosity rather than burnish a public profile.

Outside of his own indifference to the attention, it's easy to take the information revolution for granted—to simply accept that the world is as it was destined to be. That goes double for contributions like information theory, which, even to the initiated, can be difficult to understand. The reason we wrote the book is because we thought there was something wrong about that: here we've been given this bounty of digital information and instant communication, but we only have a vague appreciation for how we got here and who laid the theoretical groundwork for the world we now live in. Hopefully the book will help us all appreciate Shannon—and many of his contemporaries—a bit more.

The Historian’s Column

Anthony Ephremides

This is it! It is time to say loud and clear that the King has no clothes! I am using this metaphor to address an issue that has been enveloping our profession (and not only). The proliferation of conferences and journals that we have been experiencing over several years (if not decades) already, and which we have been ignoring and/or tolerating, has reached the point where it needs to be confronted. Or, to use another cliché, we cannot ignore anymore the elephant in the room.

It used to be that scientific gatherings and publications were serious affairs that were organized by reputable organizations and were indeed serving the needs for communication amongst scientists and engineers. To be sure, the level of activity in numerous fields, and especially in the areas that deal with communication, computation, control, and information, has been soaring and, hence, it is reasonable to expect that there would be more workshops, conferences, and symposia and more journals; especially so, since activity in these fields is no more confined to North America and Europe. In many parts of the world the needs for harnessing the creative talents and resources of local populations have yielded genuinely productive and constructive activities in intellectual and scientific discourse that has translated into more (and bigger) meetings and more periodicals and books. The freedom from the constraints that the use of traditional “print” media used to entail has led to massive use of web-based formats and has accelerated this growth.

Like any activity that acquires momentum, this one has attracted the attention of business oriented outfits (publishing houses, international meeting organizers, and creative new format inventors) and has been slowly altering the landscape we have been used to. What is worse, for the younger amongst us, there has been no alternative experience and the emerging new ecosystem is taken as the accepted status quo. And what is worst, the crowd attracted by the new opportunities of abuse and corruption in the existing system has included outright fraudsters who have grown from being a nuisance or a subject of jokes to becoming a serious threat to the foundations of the practice of science.

Every day we are bombarded with announcement of new meetings and symposia that clearly smack of fraud or, at least, degraded imitations of erstwhile serious activities. Many of us receive “invitations” to organize sessions an ANY topic we choose in conferences with ambiguous names and organized by outfits of dubious stature. The reviewing process which is supposed to provide a filter for preserving quality through the seal of approval by reputable and recognized experts has been diluted to the point of becoming meaningless. I am serving as Editor-in-Chief of two publications (more on that later) and I am in disbelief when I receive messages that in essence say something like “I am “so-and-so” and I am delighted to report to you that I am ready to edit a special issue in your esteemed journal”! Not only, in most cases, haven’t I ever heard of these individuals but sometimes they dare to include their resumes which are of unbelievably unacceptable nature. It seems that we have hit a rapid slope that threatens to lead us to a world of “alternative scientific facts” or, worse, “fake scientific news.”

Commercial publishers do share a great deal of responsibility for these developments. At best, they are guided simply by the lure of the bottom-line and they place quality at a secondary level. Beyond them, there is also a growing set of individuals and other organizations that have been polluting the international scientific scene. Without naming them, I can report that there is an individual, with claimed positions between a northern Swedish University and a University in the Persian Gulf, who has an h-index in the thousands! And there has been an organization in Europe, operating out of northern Italy, that has been running dozens (if not hundreds) of “fake” conferences after a carefully planned exploitation of originally legitimate channels of Universities and official European funding agencies.

Friends, this is dangerous and needs to be controlled. Of course we are blessed to belong to an organization whose symposia, workshops, and publications do remain at a level of the highest standard. Nonetheless, as most of us are also active in areas outside mainstream Information theory, we are undoubtedly aware of this emerging phenomenon. Perhaps not all of us have encountered cases of egregious abuse. However, most of us do perceive, I believe, a dangerous trend that is changing the landscape.

For balance, I would like to acknowledge that not all new and unusual activities are deplorable. Among the many exceptions, I would like to mention some that I happen to be involved in (not to imply that they are worthwhile because of my involvement but, rather, that I am involved because they are worthwhile). One is the Journal of Communications and Networks that has been actually on the “scene” for many years. It emerged out of the desire of the Korean community, active in the areas of communications and networking, to harness the talents of their members to lead a world-class publication that would meet the standards of high international quality. The main mover and shaker of this effort was Byung Lee, a leader in the Korea Information and Communication Society who saw that the way to establish a high quality venture was to gain the support of an organization like the IEEE. As a result of a long and sustained effort by him and others who worked with him, he managed to elevate the status of this journal through a technical sponsorship by the IEEE Communication Society and through Editorial Boards that included people like Ray Pickholtz, Steve Weinstein, Ezio Biglieri, Vince Poor, and several other well-respected members of our community,

Another one is the NOW publishing company, based in the Netherlands, which aspired to replace Springer Verlag as the brand-name of quality in assembling monographs of exceptional standard. Its endeavors include a set of Series of periodical monographs (not a contradiction of terms) called “Foundations and Trends in X” where X takes the values “Information Theory,” “Networking,” “Signal Processing,” and several others. Many of our readers are familiar with these series since NOW is usually present, and exhibits its products, during our major conference activities.

Finally, I would also like to put a plug for some conferences that might appear of questionable motivation and quality, but for which there are strong supportive arguments that are accompanied by careful implementation. One of these, in which I was recently involved, is the BALKANCOM, a small new conference on Communications and Networking that aspires to harness the talents of a developing area of the world which includes the states of the Balkan Peninsula. By the time you read this, its first edition will have taken place in Tirana, Albania with the participation of people like Gerhard Kramer, Alexandre Proutiere, George Giannakis, Petar Popovski, and other well-known members of our community. Without any sponsor, totally self-sustained, and not-for-profit this is a regional effort to establish a model for worthwhile activities that meet solid standards of quality. I hope that its future will be similar to that of the WiOpt conference that Eitan Altman and I founded in 2003 and which has been thriving and enjoying respect and international recognition ever since.

Our community needs to engage in a vigorous effort to monitor and maintain the quality that has made our Society so profoundly respected, as we engage in parallel activities; these activities occur amidst the cataclysmic environment of a veritable jungle of symposia and products that threaten the integrity of our profession.

Students' Corner: The Information Theory Student Subcommittee Goes Social!

Mine Alsan (minealsan@gmail.com) and Basak Güler (basak@psu.edu)

In order to facilitate interaction and networking opportunities between our student and postdoctoral members, the Information Theory Student Subcommittee is now present on two online platforms: Facebook and Slack.

The Facebook group was created in May 2016 with the goal to create a social media domain to connect students and postdoctoral scholars interested in information theory. The group is called the "IEEE Information Theory Society Student Group," which is a public group and can be found easily by searching its name on Facebook. Members of this group can connect with each other and learn about important news and events targeted at the student members of the information theory community, including student events in upcoming conferences as well as travel support deadlines. Interested students can send a membership request using a Facebook account to join the group.

In addition to the Facebook page, a Slack group with the name "ISIT 20**: What's up, Docs?" was initiated to create a conference-wide channel which connects the student and postdoctoral attendees of the main annual conference in our field. Slack is a communication platform exclusive for "teams" where members can post messages or share files within topic-specific channels. Attendees of ISIT can use this communication channel (which goes with a smartphone app) to engage more effectively with their fellow colleagues before, during, and after the conference. Just to highlight a few possible use cases:

- Have you ever found yourself in a situation where you needed to find a shared accommodation to attend a conference, potentially with a person outside your own institute, but did not know who to ask? Next time, feel free to drop a message to the #logistics channel to quickly reach out to your colleagues.
- Have you ever found yourself in a situation where you wished you could have presented your work to a broader audience during a conference? You can now advertise your talk to your peers via the channel #come-and-see-my-talk.
- Have you ever wished there was a section on Quora for asking one of the many questions swirling in your head at the intersection of IT, research, academia, publications, etc., but thought that no one would probably be able to generate a useful answer? In 2016, a set of more senior IT Society members kindly agreed to answer questions and give career advice in #ask-me-anything. Use this opportunity to reduce (or increase!) your uncertainties.
- Have you ever found yourself in a situation you wanted to get more out of a conference, be more productive, and get in touch with more colleagues to discuss their research interests, but did not get the opportunity in the midst of the many parallel sessions? As a first step, you can now start a conversation in #post-an-open-problem or #tutorial-materials. Alternatively, you can use the application to create a separate channel to initiate a discussion group on a topic of your choice and even organize a reading group around the conference dates.
- Have you ever found yourself in a situation where you wished there were a couple of activity buddies among the other attendees who share similar interests with you? If you don't want to miss out on the opportunity to enjoy what the conference location has to offer, we suggest that you create your own #activity channel to find like-minded people to organize group activities. In 2016, #run-a-centennial-run and #isit-bouldering were created to organize running and bouldering activities. In addition, local suggestions for restaurants around the conference venue were shared in #dinner-meet-ups. Take a holistic view during the next ISIT; you won't regret it!

In a nutshell, the goal of the Slack group is to provide the younger members of our community with a practical tool to manage more easily, shape more creatively, and enjoy to the fullest their conference experience. To join this group, simply send a private message to either Mine Alsan (minealsan@gmail.com) or Bernhard Geiger (geiger@ieee.org) with your name, surname, affiliation, and email address, and we will invite you!

Other than going social on Facebook and Slack, the IT Student Subcommittee continues to organize exciting events for IT students. For example, at ITA 2017, the Information Theory Student and Outreach Committees, together with the NSF Center for Science of Information, co-hosted the lunchtime panel "You and Your Research," which discussed Richard Hamming's renowned lecture on how to do great work. In his lecture, Hamming raises questions like: How do you manage to be a better researcher and be more successful? How can you become more productive even in a non-ideal environment? What are the most important characteristics of a good researcher? Serving as panelists, Professors Dan Costello, Michelle Effros, Emina Soljanin and Emre Telatar answered these thought-provoking questions and more.

Moreover, as in previous years, we are organizing our annual "Meet the Shannon Awardee" event at ISIT 2017. This is a great opportunity for students to learn more about this year's Shannon Award Winner Professor David Tse. The event will take place over lunchtime and students will be provided with free lunch. Last year's event with Professor Alexander Holevo can be viewed by following the link https://vimeo.com/185470703

From the Editor

continued from page 2

With sadness, we conclude this issue with a tribute to Mary Elizabeth (Betty) Moore Shannon who passed away on May 1st at the age of 95. The tribute appeared in the Boston Globe and is reproduced here with permission.

Please help to make the newsletter as interesting and informative as possible by sharing with me any ideas, initiatives, or potential newsletter contributions you may have in mind. I am in the process of searching for contributions outside our community, which may introduce our readers to new and exciting problems and, as such, broaden the influence of our society. Any ideas along this line will also be very welcome.

Announcements, news, and events intended for both the printed newsletter and the website, such as award announcements, calls for nominations, and upcoming conferences, can be submitted at the IT Society website http://www.itsoc.org. Articles and columns can be e-mailed to me at mikel@buffalo.edu with a subject line that includes the words "IT newsletter."

The next few deadlines are:

July 10, 2016 for the issue of Sep. 2017.

Oct. 10, 2016 for the issue of Dec. 2017.

Please submit plain text, LaTeX, or Word source files; do not worry about fonts or layout as this will be taken care of by IEEE layout specialists. Electronic photos and graphics should be in high resolution and sent as separate files.

I look forward to hearing your suggestions and contributions.

With best wishes,
Michael Langberg.
mikel@buffalo.edu

Thinking About Organizing an ITW or ISIT?

Emanuele Viterbo and Elza Erkip

One of the most rewarding experiences as volunteer in our Society is the organization of a workshop or conference. In this short note we would like to give some basic information and tips on how to approach this seemingly daunting task.

The team and key roles—To start, as the promoter you should try to put together a small team of volunteers possibly including someone with some prior experience, someone local (or as close as possible) to the venue, someone thinking about the technical program. Consider diversity and gender balance. The key roles that should be identified as early as possible are: General Chair (typically 1–2 persons), TPC Chair (typically 2–4 persons), Local Arrangements, Finance Chair, Publications. Other roles are also very important but can be assigned at a later stage.

The timeline—BoG meetings are scheduled three times a year around February, June/July (at the ISIT), and October. The final selection for the next available ISIT is made only at the June/July BoG meeting. ITW's can be selected at any BoG meeting. In order to prepare your bid for an ITW or ISIT you will need to submit an expression of interest to the Conference committee chair at least one to two months before the BoG meeting preceding the one where the selection is to be made (for an ISIT the expression of interest is due in early January).

The location—The selection of the location within a country should be made in close contact with the local organizers. Many factors should be considered for both ITW and ISIT: easy accessibility from main hubs, accommodation availability for different price levels (including student accommodation where possible).

The venue—The venue can be a hotel with conference facilities, a conference center or a university campus. The venue should be checked to verify the quality of the rooms and AV equipment. The plenary space should be sufficient to accommodate all the participants with a good margin. A desirable features for the venue is to have all the rooms in short range to facilitate switching sessions, and single space for coffee breaks where people can easily meet. Additional, break-out spaces for participants to sit and discuss are also useful.

The budget—A simplified budget with the key items template can be used to get an idea of the registration fees. Please contact the conference committee for a copy of this template.

The financial co-sponsorship (FCS)—The Information Theory Society is the default financial co-sponsor with IEEE of ITW and ISIT's taking responsibility for all gains and losses.

The technical co-sponsorship (TCS)—The Information Theory Society is the default technical co-sponsor of ITW and ISIT's.

The Conference Committee—The IT Society's Conference committee is in charge of collecting the expressions of interest from the promoters, give advise and feedback for the preparation of the final bid.

Resources—The most useful resource is the website of the previous editions of the conference. Previous organizers are an invaluable source of information and full manual for organizers is available from the Conference Committee. If you have further questions please contact any one of the members of the Conference Committee (http://www.itsoc.org/people/committees/conferences):

- Emanuele Viterbo, Monash University, Melbourne (Chair), emanuele.viterbo@monash.edu
- Ruediger Urbanke, EPFL, Lausanne (Ex Officio), ruediger.urbanke@epfl.ch
- Elza Erkip, New York University, (Ex Officio), elza@nyu.edu
- Daniela Tuninetti, University of Illinois at Chicago (Ex Officio), danielat@uic.edu
- Albert Guillen i Fabregas, Univ. Pompeu Fabra, Barcelona, guillen@ieee.org
- Urbashi Mitra, University of Southern California, ubli@usc.edu
- Brian Kurkoski, Japan Advanced Institute of Science and Technology, kurkoski@ieee.org
- Alfonso Martinez, Univ. Pompeu Fabra, Barcelona, alfonso.martinez@ieee.org

GOLOMB'S PUZZLE COLUMN™ COLLECTION, Part 4

Beyond his extraordinary scholarly contributions, Sol Golomb was a long time newsletter contributor enlightening us all, young and old, with his beautiful puzzles. In honor of Sol's immense contribution to the newsletter, a collection of his earlier puzzles dated back to 2001 appears in 4 compiled parts over the previous and current issues. Part 4 is given below. He will be greatly missed.

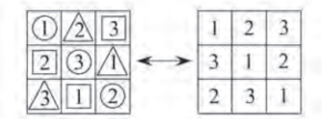
Reprinted from Vol. 53, No. 3, September 2003 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Latin Squares and Transversals Solutions

1. Given a pair of orthogonal Latin Squares, L and L', of order n, each *symbol* in L' occurs in the positions which form a transversal in L. Thus, the n symbols in L' specify n disjoint transversals in L. Conversely, if L has n disjoint transversals, each transversal of L can be used to correspond to a different symbol in L'.

For example:



2. If L is the Cayley table of a group G of order n, and it has a transversal, we can represent this as follows. Let $G = \{ g_1, g_2, \ldots, g_n \}$, and index the rows of L with $g_1, g_2, \ldots, g_n$. For a transversal in L, each row g_i must be paired with a column h_j to get an entry $g_i \times h_j = t_k$, where "×" is the group operation, and all three of $\{g_1, g_2, \ldots, g_n\}$, $\{h_1, h_2, \ldots, h_n\}$, and $\{t_1, t_2, \ldots, t_n\}$ are permutations of the n elements of G. If now we right-multiply each equation $g_i \times h_j = t_k$ by a fixed element $p \in G$, we get a "new" transversal (truly new if p is not the identity element in G), because

$$\begin{array}{c|c} g_1 \times (h_1 \times p) = (t_1 \times p) & g_1 \times h'_1 = t'_1 \\ g_2 \times (h_2 \times p) = (t_2 \times p) & g_2 \times h'_2 = t'_2 \\ \vdots & \\ g_n \times (h_n \times p) = (t_n \times p) & g_n \times h'_n = t'_n \end{array}$$

where $\{ h'_1, h'_2, \ldots, h'_n\}$ is a new permutation of the elements of G, and $\{ t'_1, t'_2, \ldots, t'_n\}$ is a new permutation of the elements of G, disjoint respectively from $\{ h_1, h_2, \ldots, h_n \}$ and $\{ t_1, t_2, \ldots, t_n \}$. Thus, each group element $p \in G$ generates a new transversal, disjoint from the others. (We used the associative law for groups when we took $(g_i \times h_j) \times p = g_i \times (h_j \times p)$.) Thus, if L is a Cayley table, one transversal gives a "complete set" of n disjoint transversals, and hence an "orthogonal mate" L'.

The converse is trivial. If L has an "orthogonal mate" L', it has n disjoint transversals, so surely at least one transversal.

3. If $p = n + 1$ is prime, $n > 1$, and L is the Cayley table of Z_p^x, the multiplicative group modulo p (which has order n), we will assume that L has a transversal, and obtain a contradiction. As in the previous problem, a transversal of L looks like:

$$(*) \qquad \begin{array}{c} g_1 \times h_1 = t_1 \\ g_2 \times h_2 = t_2 \\ \vdots \\ g_n \times h_n = t_n \end{array}$$

where $\{ g_1, g_2, \ldots g_n \}$, $\{ h_1, h_2, \ldots, h_n \}$, and $\{ t_1, t_2, \ldots, t_n \}$ are each permutations of $\{ 1, 2, \ldots, n \}$. By Wilson's Theorem of elementary number theory, $n! = (p–1)! \equiv –1 (\text{mod } p)$. Multiplying all n equations (*) together modulo p, we get

$$\begin{array}{r} (g_1 \times h_1)(g_2 \times h_2)\cdots(g_n \times h_n) \equiv t_1 \times t_2 \times \cdots \times t_n \, (\text{mod } p) \\ (p-1)! \cdot (p-1)! \equiv (p-1)! (\text{mod } p) \\ (-1) \cdot (-1) \equiv -1 (\text{mod } p) \\ +1 \equiv -1 (\text{mod } p) \end{array}$$

a contradiction.

4. "The number of mutually (pair-wise) orthogonal Latin Squares (MOLS) of order n cannot exceed $n - 1$".

Proof. It is no loss of generality to rename the elements in each of the Latin Squares so that each top row consists of 1, 2, 3, $\cdots$, n. Next, we simultaneously permute the remaining rows (which disturbs neither Latin-ness nor orthogonality) so that the second row of the first Latin Square begins with "2". We now ask what the possibilities are for the left-most elements of the second rows of the remaining, mutually orthogonal Latin Squares. From the corresponding elements in the top rows, all the ordered pairs 11, 22, 33, . . . , nn have already occurred, so these left-most positions in the second rows must all be distinct from each other, and from the "2" in the first Latin Square. Moreover, since each sits below a "1" in the top row of its Latin Square, no entry in row 2, column 1, can be "1". This leaves the n–2 values 3, 4, . . . , n, and limits the number of mutually orthogonal Latin Squares of order n to at most $1+(n - 2) = n - 1$.

Note: A construction for $n - 1$ MOLS of order n is known whenever there is a field of n elements (thus, $n = p^k$, p prime and $k \geq 1$). Whether this can happen for any other values of n is unknown in general, has been shown to be impossible for infinitely many $n \neq p^k$, and has never been successfully constructed for any $n \neq p^k$.

5. "If a Latin Square L of order n has n–1 disjoint transversals, then it has n disjoint transversals."

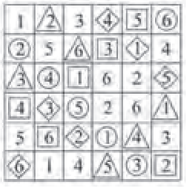
Proof. Each transversal of L consumes one cell in each row, one cell in each column, and one of each of the n symbols. Hence, $n - 1$ disjoint transversals consume $n - 1$ cells in each row, $n - 1$ cells in each column, and $n - 1$ of each of the n symbols. Thus, what remains in L is one cell in each row, one cell in each column, and one each of the n symbols, i.e. an n^{th} disjoint transversal.

6. Here is a Latin Square of order 6 with four disjoint transversals, whose elements are enclosed in the figures,circle, triangle, square, and diamond, respectively.

Note: Euler further conjectured that no pair of orthogonal Latin Squares of order n exists when $n = 2m$, where m

Reprinted from Vol. 53, No. 3, September 2003 issue of Information Theory Newsletter continued

is odd. This was shown to be false, in 1959, for all $n > 6$, by R.C. Bose, S.S. Shrikhande, and E.T. Parker. In fact, it has been shown that if $M(n)$ is the maximum number of MOLS of order n which actually occur, then $\liminf_{n\to\infty} M(n) = \infty$.



1	△2	3	◇4	□5	○6
○2	5	△6	□3	◇1	4
△3	○4	□1	6	2	◇5
□4	◇3	○5	2	6	△1
5	□6	◇2	○1	△4	3
◇6	1	4	△5	○3	□2

Reprinted from Vol. 53, No. 4, December 2003 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Irreducible Divisors of Trinomials Solutions

Solomon W. Golomb

1. "A primitive polynomial $f(x)$ of degree $n \geq 2$ divides infinitely many trinomials over $GF(2)$".

Proof. Let α be a root of $f(x)$. By "primitivity", all the values $1, \alpha, \alpha^2, \alpha^3, \ldots, \alpha^{2^n-2}$ are distinct, and are all the non-zero elements of $GF(2^n)$. Therefore, for each $j, 0 < j < 2^n - 1$, $1 + \alpha^j = \alpha^k$ with $0 < k < 2^n - 1$ and $j \neq k$. Hence, $f(x)$ divides the trinomial $1 + x^j + x^k$ for these values of j and k. In addition, $f(x)$ divides $1 + x^J + x^K$ for every J with $J \equiv j \pmod{2^n - 1}$ and $K \equiv k \pmod{2^n - 1}$, since $1 + \alpha^J + \alpha^K = 1 + \alpha^j + \alpha^k = 0$, in view of $\alpha^{2^n-1} = 1$.

2. "If $f(x)$ is irreducible with primitivity t and $f(x)$ divides no trinomials of degree $< t$, then $f(x)$ divides no trinomials."

Proof (by contradiction). Suppose $f(x)$ divides the trinomial $x^N + x^A + 1$ of degree $N > t$, and let α be a root of $f(x)$. Since $f(x)$ has primitivity t, $\alpha^t = 1$. Since $f(\alpha) = 0$, where α is a root of $f(x)$, any polynomial $g(x)$ divisible by $f(x)$ also has α as a root, since $g(x) = f(x) \cdot q(x)$ gives $g(\alpha) = f(\alpha) \cdot q(\alpha) = 0 \cdot q(\alpha) = 0$. Thus, $\alpha^N + \alpha^A + 1 = 0$, from which $\alpha^n + \alpha^a + 1 = 0$ where $n \equiv N \pmod{t}$ and $a \equiv A \pmod{t}$, where we choose both n and a to be less than t, from which $f(x)$ divides the trinomial $x^n + x^a + 1$, of degree $<t$.

3. "If $p \geq 5$ is a prime for which 2 is primitive modulo p, then $f(x) = (x^p - 1)/(x - 1) = 1 + x + x^2 + \cdots + x^{p-1}$ is an irreducible polynomial which divides no trinomials."

Proof. For each prime p, $\Phi_p(x) = (x^p - 1)/(x - 1)$ is the "cyclotomic polynomial" over the rational field Q, whose roots are the $\phi(p) = p - 1$ primitive p^{th}roots of unity. While all cyclotomic polynomials are irreducible over Q, $\Phi_p(x)$ remains irreducible over $GF(2)$ if and only if 2 is primitive modulo p. In this case, $f(x) = \Phi_p(x)$ has primitivity $t = p$, and any root α of this $f(x)$ has $\alpha^p = 1$. Note that for $p \geq 5$, the minimum polynomial for such a root of unity has $p > 3$ non-zero terms. By Result 2, above, if this $f(x)$ divides *any* trinomial, it must divide a trinomial of degree $<t = p$, say $x^n + x^a + 1$ with $n < p$. But then the root α of $f(x)$ is a root of this trinomial of degree $\leq p - 1$, whereas the unique polynomial of degree $\leq p - 1$ with α as a root is the *minimal* polynomial of α, $f(x) = \Phi_p(x)$, of degree $p - 1$,which has *more than three terms.*

Note. There are also many other irreducible polynomials which divide no trinomials. These three problems are the easy results.

Reprinted from Vol. 54, No. 1, March 2004 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

OVERLAPPING SUBSETS SOLUTIONS

Solomon W. Golomb

1. a. By statistical independence, the expected number of overlaps is $M = (\frac{a}{N})(\frac{b}{N})N = \frac{ab}{N}$.

   b. $pr(k) = \frac{\binom{a}{k}\binom{N-a}{b-k}}{\binom{N}{b}} = \frac{\binom{b}{k}\binom{N-b}{a-k}}{\binom{N}{a}} = \frac{a!b!(N-a)!(N-b)!}{k!(a-k)!(b-k)!N!(N-a-b+k)!}$.

   c. $\frac{pr(k+1)}{pr(k)} = \frac{(a-k)(b-k)}{(k+1)(N-a-b+k+1)}$.

2. a. $M = 9$.

   b.

k	$\frac{pr(k+1)}{pr(k)}$	k	$\frac{pr(k+1)}{pr(k)}$	k	$\frac{pr(k+1)}{pr(k)}$
0	11.2344	4	2.0403	8	1.0284
1	5.4855	5	1.6586	9	0.8999
2	3.5703	6	1.3865	10	0.7959
3	2.6136	7	1.1829	11	0.7105

   c. The *mode* is 9 (same as the *mean*, in this case), since $pr(k)$ is *increasing* up to $k + 1 = 9$, but *decreasing* thereafter.

3. a. $pr(k) = \frac{\binom{90}{k}\binom{810}{90-k}}{\binom{900}{90}} = \frac{(90!)^2(810!)^2}{k!((90-k)!)^2 900!(720+k)!}$

At $k = 9$,

$$pr(9) \approx \frac{(2\pi)^2 \cdot 90 \cdot 810 \cdot (\frac{90}{e})^{180} \cdot (\frac{810}{e})^{1620}}{(2\pi)^{\frac{5}{2}} \sqrt{9 \cdot 81^2 \cdot 900 \cdot 729} (\frac{9}{e})^9 (\frac{81}{e})^{162} (\frac{900}{e})^{900} (\frac{729}{e})^{729}}.$$

   b. Except for a factor of $\sqrt{2\pi}$ in the denominator, all the irrational numbers disappear. (The powers of e cancel completely between numerator and denominator. Fortuitously, 9, 81, 729, and 900 are all perfect squares; and everything surviving involves only powers of 3 and of 10.) When all the smoke clears, all that remains is $pr(9) \approx \frac{10}{27\sqrt{2\pi}}$.

   c. Numerically, $pr(9) \approx \frac{10}{27\sqrt{2\pi}} = 0.1477564$.

4. a. $Pr(9) = e^{-9} \cdot \frac{9^9}{9!} = 0.13175564$.

   b. The largest source of error in 3.c. was using Stirling's formula to approximate 9! in the denominator of 3.a., which gives $9! \approx 359, 536.873$. This is only about 99% of the true value ($9! = 362, 880$). This "correction" would only reduce the estimate in 3.c. to $pr(9) \approx 0.146$; so 3.c. is almost certainly a better estimate than 4.a.

   c. Since $\frac{Pr(k+1)}{Pr(k)} = \frac{\lambda}{k+1}$ for the Poisson distribution, if we take $\lambda = 9$ and $7 \leq k \leq 11$, we find

k	$\frac{9}{(k+1)}$
7	1.125
8	1.000
9	0.900
10	0.818
11	0.750

   which are fairly close to the values in 2.b. (The values will not be as close for k farther from λ.)

5. $Pr(25) = 5.712 \times 10^{-6}$ when $\lambda = 9$. (The true value of $pr(25)$ is about 2.2×10^{-7}, and is actually much smaller than the Poisson approximation.) The student's intuition was correct.

Reprinted from Vol. 54, No. 2, June 2004 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

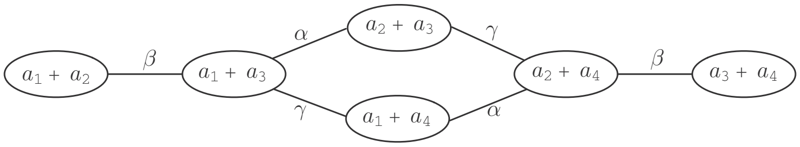
An Inverse Problem—Solutions

Solomon W. Golomb

We are asked to reconstruct a set S of n distinct positive real numbers, given only the set T consisting of the $\binom{n}{k}$ sums of the k-element subsets of S. Let the elements of S be $a_1 < a_2 < a_3 < \cdots < a_n$. Each a_i occurs in exactly $\binom{n-1}{k-1}$of the k-element subsets of S. Hence, the sum, $a_1 + a_2 + \cdots + a_n$, of all the elements of S can be obtained by summing all $\binom{n}{k}$ elements of T and then dividing by $\binom{n-1}{k-1}$.

If $n > k$, the *smallest* element of T is $a_1 + a_2 + \cdots + a_k$, and the *next-smallest* element of T is $a_1 + a_2 + \cdots + a_{k-1} + a_{k+1}$. Similarly, the *largest* element of T is $a_n + a_{n-1} + \cdots + a_{n-k+1}$, and the *next-largest* element of T is $a_n + a_{n-1} + \cdots + a_{n-k+2} + a_{n-k}$. The remaining elements of T are partially ordered by magnitude. This partial ordering can usefully be shown by a graph, where the nodes are the elements of T (increasing in numerical magnitude from left to right), and the edges are labeled with the difference of the magnitudes of the nodes they connect. The only distinct edge labels will be $\alpha = a_2 - a_1$, $\beta = a_3 - a_2$, $\gamma = a_4 - a_3$, etc. These facts, and the corresponding graphs, will be used to solve problems 1 to 4 as follows.

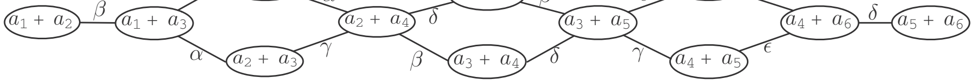
1. $n = 4$, $k = 2$, $T = \{24, 28, 30, 32, 34, 38\}$.



We know $a_1 + a_2 = 24$, $a_1 + a_3 = 28$, $a_2 + a_4 = 34$, $a_3 + a_4 = 38$, and $\beta = a_3 - a_2 = 4$. There are two possibilities, leading to two solutions. Either $a_2 + a_3 = 30$ and $a_1 + a_4 = 32$, or $a_2 + a_3 = 32$ and $a_1 + a_4 = 30$. In the former case $\alpha = 2$, $\gamma = 4$, while in the latter case $\alpha = 4$, $\gamma = 2$. In the former case, $a_3 + a_2 = 30$, $a_3 - a_2 = \beta = 4$, and $a_3 = 17$, giving $a_2 = 13$, $a_1 = 11$, and $a_4 = 21$. That is, a first solution is $S = \{11, 13, 17, 21\}$. In the latter case, $a_3 + a_2 = 32$, $a_3 - a_2 = \beta = 4$, and $a_3 = 18$, from which $a_2 = 14$, $a_1 = 10$, and $a_4 = 20$. That is, the second solution is $S = \{10, 14, 18, 20\}$.

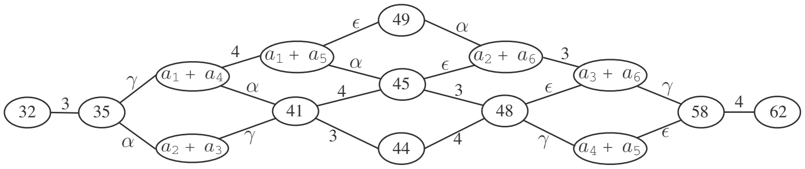
2. $n = 5$, $k = 2$, $T = \{21, 26, 28, 29, 31, 34, 36, 37, 42, 44\}$. With $S = \{a_1, a_2, a_3, a_4, a_5\}$ with $a_1 < a_2 < a_3 < a_4 < a_5$, we have $a_1 + a_2 + a_3 + a_4 + a_5 = 328/4 = 82$, where 328 is the sum of the elements in T, and $4 = \binom{5-1}{2-1}$. We know $a_1 + a_2 = 21$, $a_1 + a_3 = 26$, $a_5 + a_4 = 44$, and $a_5 + a_3 = 42$. Also, $a_3 = 82 - (a_1 + a_2) - (a_4 + a_5) = 82 - 21 - 44 = 17$. Then $a_1 = 26 - 17 = 9$ and $a_5 = 42 - 17 = 25$. Finally, $a_2 = 21 - 9 = 12$, and $a_4 = 44 - 25 = 19$. Hence the unique solution is $S = \{9, 12, 17, 19, 25\}$.

3. $n = 6$, $k = 2$, $T = \{32, 35, 37, 39, 41, 43, 44, 45, 48, 49, 51, 52, 54, 58, 62\}$. Here the graph is

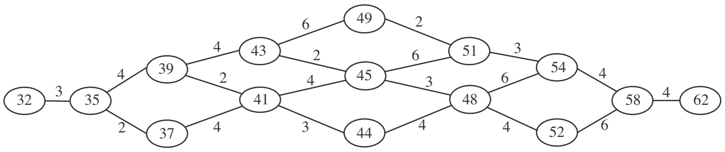


We know $a_1 + a_2 = 32$, $a_1 + a_3 = 35$, $\beta = 3$, $a_5 + a_6 = 62$, $a_4 + a_6 = 58$, $\delta = 4$. Also, $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 = 690/5 = 138$, from which $a_3 + a_4 = 138 - (a_1 + a_2) - (a_5 + a_6) = 138 - 32 - 62 = 44$; $a_2 + a_5 = 138 - (a_1 + a_3) - (a_4 + a_6) = 138 - 35 - 58 = 45$, and $a_1 + a_6 = 138 - (a_2 + a_5) - (a_3 + a_4) = 138 - 45 - 44 = 49$. Our graph now becomes

Reprinted from Vol. 54, No. 2, June 2004 issue of Information Theory Newsletter continued

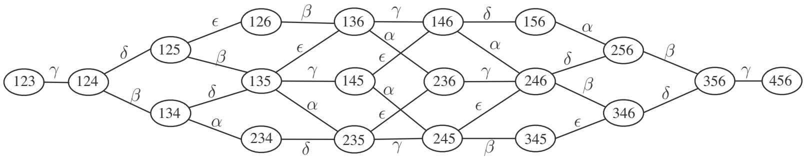


From nodes 35 to 41, $\alpha + \gamma = 6$. From nodes 48 to 58, $\gamma + \epsilon = 10$. From nodes 45 to 49, $-\alpha + \epsilon = 4$. (This is a dependent set of 3 equations, so we do not yet have a unique solution.) The third-smallest element of T, 37, is either $a_1 + a_4$ or $a_2 + a_3$, so either $\gamma = 2$ or $\alpha = 2$. The third-largest element of T, 54, is either $a_3 + a_6$ or $a_4 + a_5$, so either $\gamma = 4$ or $\epsilon = 4$. From the last two statements, there are three possibilities: i) $\gamma = 2, \epsilon = 4$; ii) $\alpha = 2, \gamma = 4$; iii) $\alpha = 2, \epsilon = 4$. Since $\gamma + \epsilon = 10$ we rule out i). Since $-\alpha + \epsilon = 4$ we rule out iii). That leaves only ii), with $\alpha + \gamma = 2 + 4 = 6$. We can now uniquely fill in the entire graph.



Knowing which two elements of S were summed to obtain each element of T, we have 15 consistent linear equations in only 6 unknowns. One easy way to solve this system: $a_3 - a_2 = \beta = 3$, $a_3 + a_2 = 37$, hence $a_3 = 20$, $a_2 = 17$, and $a_1 = 32 - 17 = 15$. Also $a_5 - a_4 = \delta = 4$, $a_5 + a_4 = 52$, $a_5 = 28$, hence $a_4 = 24$, and $a_6 = 62 - 28 = 34$. Thus, the unique solution is $S = \{15, 17, 20, 24, 28, 34\}$.

4. $n = 6, k = 3, T = \{49, 54, 56, 57, 58, 60, 61, 65, 66, 67, 68, 69, 70, 74, 75, 77, 78, 79, 81, 86\}$. We have $a_1 + a_2 + a_3 = 49$, $a_1 + a_2 + a_4 = 54$, $a_6 + a_5 + a_4 = 86$, $a_6 + a_5 + a_3 = 81$. Thus $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 = 135$, and $a_4 - a_3 = \gamma = 5$. The graph, using only the subscripts of the summed a_i's to label the nodes, shows its central symmetry, and in fact it has the V_4 symmetry group of the rectangle.



It is relatively easy to show that there are only two sets of six positive real numbers for which the sums of all 3-subsets yield the twenty elements of T. One is $S_1 = \{15, 16, 18, 23, 27, 36\}$, and the other is $S_2 = \{9, 18, 22, 27, 29, 30\}$. These correspond to two mirror-image assignments to the edges of the graph. S_1 uses $\alpha = 1, \beta = 2, \gamma = 5, \delta = 4, \epsilon = 9$, while S_2 uses the reverse assignment $\alpha = 9, \beta = 4, \gamma = 5, \delta = 2, \epsilon = 1$.

5. If $k' = n - k$, the k'-subsets of $S = \{a_1, a_2, \ldots, a_n\}$ are precisely the *complements* (relative to S) of the k-subsets. Since we showed how to obtain the sum τ of all elements in S $\left(\text{by dividing the sum of all elements of } T \text{ by } \binom{n-1}{k-1}\right)$, if we replace the elements of T by τ minus each of these elements to obtain T', we see the equivalence of the two problems.

6. For $n = k = 2$, we have $S = \{a_1, a_2\}$ and $T = \{a_1 + a_2\}$. From the single positive element in T, there are infinitely many ways (a continuum of ways) to represent it as a sum of two elements. For $n = 3, k = 2$, we have $S = \{a_1, a_2, a_3\}$ and $T = \{a_1 + a_2, a_1 + a_3, a_2 + a_3\}$. If $a_1 < a_2 < a_3$ then $a_1 + a_2 < a_1 + a_3 < a_2 + a_3$, so if we are given $T = \{r, s, t\}$ with $r < s < t$, then we have a solvable system of three linear equations in three unknowns, with a unique solution. We saw in Problem 1 that the case $n = 4, k = 2$, has *two* solutions for S, while in Problems 2 and 3 we saw that $n = 5, k = 2$, and $n = 6, k = 2$, have unique solutions. For $k = 2$ and $n > 4$ the reconstruction of S from T is unique.

7. If $n = k > 1$, then S contains at least two elements while T contains only one (the sum of all elements of S), and as in the case $n = k = 2$, there are infinitely many solutions.

8. For $k \geq 2$ and $n = 2k$ there will be two solutions for S, given T. This is the case where $k' = k$ (in Problem 5), and the graph has V_4 symmetry, whereby each solution has a complementary solution. (In Problems 1 and 4, we saw the special cases $n = 4, k = 2$, and $n = 6, k = 3$.)

Reprinted from Vol. 54, No. 3, September 2004 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

SOME PRIME NUMBER PROPERTIES—Solutions

Solomon W. Golomb

Here $p_n = n^{th}$ prime number, and $\pi(x)$ = number of primes $\leq x$, for positive real x.

1. "Prove that the ratio $\frac{n}{\pi(n)}$, for $n \geq 2$, takes every integer value > 1 at least once."
   Proof. It was given in the Puzzle Column that $\lim\limits_{x\to\infty} \frac{\pi(x)}{x} = 0$ and $\lim\limits_{n\to\infty} p_n = \infty$. Thus the ratio $\frac{\pi(x)}{x}$ ultimately becomes and remains less than any assigned $\epsilon > 0$, as $x \to \infty$. It starts at $\frac{\pi(2)}{2} = \frac{1}{2}$.
   For any $m \geq 2$, there is a *unique largest prime* $p_k = p_{k(m)}$ for which $\pi(p_k) = k \geq \frac{p_k}{m}$. Thus, $m\pi(p_k) = mk \geq p_k$. Either $mk < p_{k+1}$ or $mk \geq p_{k+1}$.
   If $mk < p_{k+1}$, and since $p_k \leq mk, \pi(p_k) \leq \pi(mk) < \pi(p_{k+1})$, from which $\pi(mk) = k$, and $\frac{mk}{\pi(mk)} = m$, so that $n = mk$ is an integer for which $\frac{n}{\pi(n)} = m$.
   If $mk \geq p_{k+1}$, then $\pi(p_{k+1}) = k + 1 > k = \frac{mk}{m} \geq \frac{p_{k+1}}{m}$, which contradicts the choice of p_k as the *largest* prime for which $\pi(p) \geq \frac{p}{m}$. . □

2. "Every positive integer belongs to exactly one of the two sequences $\{s_n\} = \{n + \pi(n)\}$ and $\{t_n\} = \{n + p_n - 1\}$."
   Proof. In the land of Primordia, the sequence $\{p_n\}$ is used as a "tax table", in the sense that the sales tax increases by one cent at every term of the sequence $\{p_n\}$ (and at no other values). Thus the sales tax on the price p_k is exactly k. More generally, the sales tax on the price m is $\pi(m)$, the number of terms of $\{p_n\}$ not exceeding m.
   From this point of view, the "total price" (including tax) on an item with a net price of n is $n + \pi(n)$. The sequence $\{n + \pi(n)\}$ thus consists of all numbers which can occur as "total prices". What numbers cannot occur as "total prices"? As the net price increases through one of the terms of $\{p_n\}$, say from $p_n - 1$ to p_n, the total price increases from $(p_n - 1) + (n - 1)$ to $p_n + n$, thus skipping the value $p_n + n - 1$. If m is not of the form p_n, then the total price goes from $(m - 1) + \pi(m - 1)$ to $m + \pi(m)$, increasing by only one cent, because in this case $\pi(m - 1) = \pi(m)$. Thus the integers skipped in the sequence $\{n + \pi(n)\}$ are precisely the terms of the sequence $\{n + p_n - 1\}$. □

Note that in problems 1 and 2, the fact that $\{p_n\}$ is the sequence of the prime numbers (rather than some other subsequence of the positive integers that becomes less dense) plays almost no role.

3. "Given positive integers a and b, there exists a positive integer c such that infinitely many numbers of the form $an + b$ (n a positive integer) have all their prime factors $\leq c$."
   Proof. All numbers in the sequence $\{b(a + 1)^k, k = 1, 2, 3, \ldots\}$ are distinct and of the form $an + b$. Thus $c = \max(b, a + 1)$ satisfies the condition of the problem. □

4. (a) "What is the largest integer N such that, if $1 < k < N$ and k has no prime factor in common with N, then k is prime?"
   Answer. $N = 30$. Since $30 = 2 \times 3 \times 5 = p_1 \times p_2 \times p_3$, the product of the first three primes, every k relatively prime to 30 cannot be divisible by 2 or 3 or 5, and the smallest $k > 1$ divisible by none of these and *not* prime is $p_4^2 = 7^2 = 49$, which is bigger than 30.

   (b) "What is the largest *odd* integer N such that, if $1 < k < N$ and k has no prime factor in common with $2N$, then k is prime?"
   Answer. $N = 105$. Since $105 = 3 \times 5 \times 7 = p_2 \times p_3 \times p_4$, the product of the first three *odd* primes, every *odd* k relatively prime to 105 (i.e. every k relatively prime to $2 \times 105 = 210$) cannot be divisible by 2 or 3 or 5 or 7, and the smallest k divisible by none of these and *not* prime is $p_5^2 = 11^2 = 121$, which is bigger than 105.

(Known results on the distribution of the primes prevent larger solutions than 30 and 105 to these problems.)

5. "For what positive integers n is it true that $\sum\limits_{p\leq\pi(n)} p = n$?"

Answer. $n = \{5, 17, 41, 77, 100\}$. For ”large” n, $\sum\limits_{p \le \pi(n)} p > n$, and $n = 100$ is the last value for which equality holds. □

(For a more detailed solution, see the Solution to Problem E3385 (**American Math. Monthly**) listed as Reference 5 at the end of these Solutions.)

6. Let $a_1 < a_2 < a_3 < \cdots$ be an increasing, infinite sequence of positive integers.”

   (a) “Construct such a sequence $\{a_k\}$ having the property that, for *every* integer n (positive, negative, or zero) the sequence $\{a_k + n\}$ contains only finitely many prime numbers.”
   Construction. Let $\{a_k\} = \left\{((2k)!)^3\right\}$ for $k = 1, 2, 3, \ldots$. For $A_n = \{a_k + n\}$, if $|n| \ge 2$ then all terms of A_n with $k \ge n$ are divisible by n, and hence not prime. For $n = 0$, $A_n = \{a_k\}$ is clearly composite for *all* $k \ge 1$. Finally, using $x^3 + 1 = (x + 1)(x^2 - x + 1)$ and $x^3 - 1 = (x - 1)(x^2 + x + 1)$, the values of A_n for $n = \pm 1$ are composite for all $k \ge 2$.
   (Many other constructions are possible.)

   (b) “Is there such a sequence $\{a_k\}$ and a constant $B > 0$ such that, for every integer n (positive, negative, or zero) the sequence $A_n = \{a_k + n\}$ contains no more than B prime numbers?”
   The answer to this is unknown. A “yes” answer would contradict the “prime k-tuples” conjecture, which the late Paul Erdös was convinced had to be true. However, at least two other plausible conjectures in prime number theory also contradict the “prime k-tuples” conjecture.
   On the model of the construction given in 6.(a) above, let $\{a_k\} = \{((2k)!)^{3F(k)}\}$, where $F(k)$ can be an integer-valued function that grows uncomputably fast. Then each translate sequence $A_n = \{a_k + n\}$ will be “expected” (by the thinning density of the sequence of prime numbers) to contain only a *small* finite number of prime numbers, whereas our uniform bound B can be chosen arbitrarily large (e.g. $B = 10^{10^{10^{100}}}$). Anyone who can exhibit a *specific* sequence $\{a_k\}$ such that all its translates *provably* contain no more than B primes each (for a specific B, however large) will earn a permanent place in the history of prime number theory.

All the problems in this set are based on articles or problems which I published (over several decades) in either the **American Mathematical Monthly (AMM) or in Mathematics Magazine (Math. Mag.)**.

1. “On the ratio of N to $\pi(N)$”, **AMM**, vol. 69, no. 1, Jan. 1962, 36-37.

2. “The ‘Sales Tax’ Theorem”, **Math. Mag.**, vol. 49, no. 4, Sep.-Oct. 1976, 187-189.

3. Problem E2725, **AMM**, vol. 85, no. 7, Aug.-Sep., 1978, p. 593. Solution, vol. 86, no. 9, November, 1979, p. 790.

4. Problem E3137, **AMM**, vol. 93, no. 3, March, 1986, p. 215. Solution, vol. 94, no. 9, November, 1987, p. 883.

5. Problem E3385, **AMM**, vol. 97, no. 5, May, 1990, p. 427. Solution, vol. 98, no. 9, November, 1991, pp. 858-859.

6. Problem 10208, **AMM**, vol. 99, no. 3, March, 1992, p. 266. Solution, vol. 102, no. 4, April, 1995, pp. 361-362.

For further information about Problem 6.(b) and its relation to the “prime k-tuples conjecture”, see Paolo Ribenboim, *The Little Book of Bigger Primes*, Second Edition, Springer, New York, 2004, pp. 201-204, where the existence of a solution to 6.(b) is called “Golomb’s Conjecture”.

Reprinted from Vol. 54, No. 4, December 2004 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Countable or Uncountable Solutions

Solomon W. Golomb

In all three problems, $S = \{A_i\}$ is a collection of infinite subsets (or, infinite subsequences) A_i of the positive integers.

1. If $A_i \cap A_j = \emptyset$ whenever $i \neq j$, then S is (at most) countably infinite.

   Proof. Let M be the collection of smallest elements of all the sets A_i. Since the sets A_i are pairwise disjoint, each one contributes a *different* positive integer to M; so M is (at most) countably infinite; but the elements of M are in one-to-one correspondence with the elements A_i of S. □

2. If $A_i \cap A_j$ is finite (or empty) whenever $i \neq j$, it is possible for S to be uncountably infinite. Here is one such construction. For each real number α on the interval $(\frac{1}{2}, 1)$, write the binary expansion of α in the form $\alpha = 0.1a_2a_3a_4\ldots$, and associate with α the subsequence A_α of the positive integers consisting of $\{1, 1a_2, 1a_2a_3, 1a_2a_3a_4, \ldots\}$ where these are the binary representations of integers. (For example, $\alpha = \frac{2}{3} = 0.1010101\ldots$ is associated with the sequences $\{1, 10, 101, 1010, 10101, \ldots\}$ of integers in binary representation, or $\{1, 2, 5, 10, 21, \ldots\}$ in decimal notation.) For those real numbers with two representations, one "terminating" and one "repeating", we can use either representation; for specificity, let us use the repeating representation. (For example, $\alpha = \frac{3}{4}$ can be written as either 0.1100000... or 0.101111111.... For the former representation the sequence becomes $\{1, 3, 6, 12, 24, 48, \ldots\}$; for the latter representation, it becomes $\{1, 2, 5, 11, 23, 47, \ldots\}$. For the present construction, we can use either of these; and in fact the argument is *strengthened* if we use *both*.) Suppose $\beta \neq \alpha$ where $\beta = 0.1b_2b_3b_4b_5\ldots$ is the binary expansion of β. The *sequence* A_β has only finitely many terms in common with the sequence A_α; because, since $\alpha \neq \beta$, there is a smallest t for which $a_t \neq b_t$. Then not only $1a_2a_3\cdots a_t \neq 1b_2b_3\cdots b_t$, but all *subsequent* integers in A_α and A_β are different. Thus, $A_\alpha \cap A_\beta$ is a finite set for every $\alpha \neq \beta$, and there is such a set A_α for every α in the uncountably infinite set of real numbers in the interval $(\frac{1}{2}, 1)$. □

3. If $A_i \cap A_j$ has at most m elements whenever $i \neq j$, then S is (at most) countably infinite.

   Proof. Replace each set A_i by the finite set F_i consisting of the $m+1$ smallest elements of A_i. Then if $A_i \neq A_j$ we must have $F_i \neq F_j$, because A_i and A_j can have at most m common elements, by hypothesis. Hence there is a one-to-one correspondence between the sets A_i and the sets F_i. But the collection of *all* finite subsets of the positive integers is a countably infinite collection, and a fortiori the collection of subsets of $m+1$ elements from the set of positive integers is a countably infinite collection; so $S = \{A_i\}$ is (at most) a countably infinite collection. □

Note the similarity of the solutions given here for Problem 1 and Problem 3.

Reprinted from Vol. 55, No. 1, March 2005 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

A Quadratic Sequence Solutions

Solomon W. Golomb

The problems concern the sequence $S = \{s_n\} = \{2n^2 + 2n + 1\}$. Note that $s_n = 2n^2 + 2n + 1 = n^2 + (n+1)^2 = ((2n+1)^2 + 1)/2$. In particular, each s_n is a sum of two consecutive squares. The following facts are well-known from elementary number theory:

a. The primes which are sums of two squares are those of the form $4m + 1$, and 2.

b. If u and v are relatively prime, all prime factors of $u^2 + v^2$ are primes which are sums of two squares. If further $u^2 + v^2$ is odd, its prime factors must all be of the form $4m + 1$.

c. For primes p of the form $4m + 1$, the number -1 is a *quadratic residue* modulo p. That is, there is a number a such that $a^2 \equiv -1 (\text{mod } p)$; and then also $b = p - a$ satisfies $b^2 \equiv -1 (\text{mod } p)$.

Now for the solutions.

1. Since n and $n + 1$ are relatively prime, and $s_n = n^2 + (n+1)^2$ is odd, all prime factors of s_n are primes of the form $4m + 1$, for every s_n.

2. and 3. Given any $p = 4m + 1$, a prime, we will find values of n such that p divides $s_n = ((2n+1)^2 + 1)/2$, i.e., such that $((2n+1)^2 + 1)/2 \equiv 0 (\text{mod } p)$.Multiplying both sides by 2, this says $(2n+1)^2 \equiv -1 (\text{mod } p)$, and by fact c., there are numbers a_0 and $d_0 = p - a_0$ with $a_0^2 \equiv d_0^2 \equiv -1 (\text{mod } p)$. Since $a_0 + d_0 = p$, which is odd, one of a_0 and d_0 must be odd, say a_0. Then take $a_0 = 2n + 1$, so that $n = (a_0 - 1)/2$.For this value of n,and all other n congruent to it modulo p,s_n is a multiple of p, i.e., $2a^2 + 2a + 1$ is a multiple of p for all $a \equiv a_0 (\text{mod } p)$; i.e., all $a = a_0 + kp$ for all integers k. Now take $b_0 = p - a_0 - 1$. For $n = b_0$, $s_n = 2b_0^2 + 2b_0 + 1 = 2(p - a_0 - 1)^2 + 2(p - a_0 - 1) + 1 \equiv 2a_0^2 + 4a_0 + 2 - 2a_0 - 2 + 1 \equiv 2a_0^2 + 2a_0 + 1 \equiv 0 (\text{mod } p)$. Thus p divides s_n for $n = b_0$ and for all $n \equiv b_0 (\text{mod } p)$. Note also that $a_0 \neq b_0$, for otherwise $a_0 = b_0 = \frac{p-1}{2}$, leading to $a_0^2 \equiv b_0^2 \equiv (\frac{p-1}{2})^2 \equiv -1 (\text{mod } p)$, from which $(p-1)^2 \equiv -4 (\text{mod } p)$; but $(p-1)^2 \equiv (-1)^2 \equiv 1 \equiv -4 (\text{mod } p)$, and $5 \equiv 0 (\text{mod } p)$, which happens only for $p = 5$. However, looking at the actual sequence S, s_n is not divisible by 5 when $n = 2 = \frac{5-1}{2}$, but when $n = a = 1$, and when $n = b = 5 - 1 - 1 = 3$.

4. For $s_n = n^2 + (n+1)^2 = c^2$, we have a "Pythagorean triple" of the special form $(n, n+1, c)$, such as (3, 4, 5) and (20, 21, 29). We find the general solution as follows: Suppose $s_n = ((2n+1)^2 + 1)/2 = y^2$ for some y. Set $x = 2n + 1$, so that $x^2 + 1 = 2y^2$, $x^2 - 2y^2 = -1$.This is a case of "Pell's equation", which we now solve. The s*mallest* solution has $x_0 = y_0 = 1$, so that $1^2 - 2 \cdot 1^2 = -1$. We factor the Pell equation to get $(x + \sqrt{2}y)(x - \sqrt{2}y) = -1$, and then $(x + \sqrt{2}y)^n (x - \sqrt{2}y)^n = (-1)^n$. At $n = 2$, we have $(x_1 + \sqrt{2}y_1)^2 = (1 + \sqrt{2})^2 = 3 + 2\sqrt{2}$, for $x_2 = 3, y_2 = 2$, as in $3^2 - 2 \cdot 2^2 = (-1)^2 = +1$. At $n = 3$, $(1 + \sqrt{2})^3 = 7 + 5\sqrt{2}$,corresponding to $7^2 - 2 \cdot 5^2 = (-1)^3 = -1$.It is only for *odd* values of n that we get $x_n^2 - 2y_n^2 = -1$. Remember that $x = 2n + 1$, so that $n = (x - 1)/2$. From $x_1 = 1, n_1 = 0$, and $s_0 = 0^2 + 1^2 = 1^2$. From $x_3 = 7, n_3 = 3$, and $s_3 = 3^2 + 4^2 = 5^2$. At $n = 4$, we have $(1 + \sqrt{2})^4 = 17 + 12\sqrt{2}$, as in $17^2 - 2 \cdot 12^2 = (-1)^4 = +1$. At $n = 5$, we have $(1 + \sqrt{2})^5 = 41 + 29\sqrt{2}$. From $x = 41, n = 20$, and $s_{20} = 20^2 + 21^2 = 841 = 29^2$. Next, $(1 + \sqrt{2})^6 = 99 + 7\sqrt{2}$, where $99^2 - 2 \cdot 70^2 = +1$; but $(1 + \sqrt{2})^7 = 239 + 169\sqrt{2}$, where $239^2 - 2 \cdot 169^2 = -1$. With $x = 239$, $n_5 = 119$, and $s_{119} = 119^2 + 120^2 = 28,561 = 169^2$.

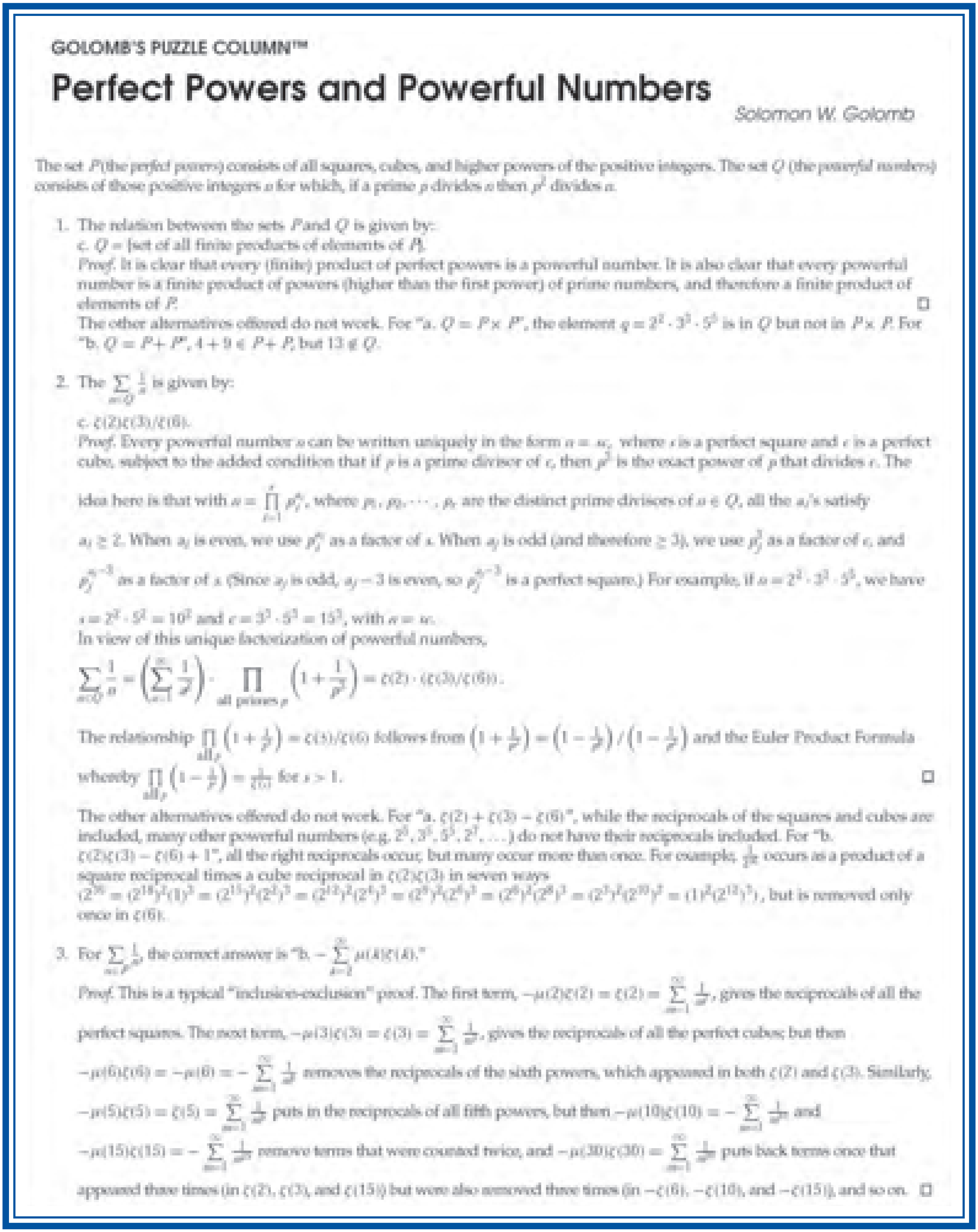
5. Thus, the sub-sequence $C = \{c_1, c_2, c_3, c_4, \ldots\}$ of *square roots* of the squares in sequence S begins $C = \{1, 5, 29, 169, \ldots\}$, and can easily be generated recursively by: $c_0 = 1, c_{n+1} = 6c_n - c_{n-1}$ for $n \geq 1$. (This can be proved inductively from the Pell equation approach. It yields a much simpler way of obtaining the values of the c_n's.) Thus $c = \{1, 5, 29, 169, 985, 5741, 33461, 195025, 1136689, 6625109, 38613965, 225058681, 1311738121, 7645370045, 44560482149, 259717522849, \ldots\}$. The recursion $c_{n+1} = 6c_n - c_{n-1}$ corresponds to the polynomial equation $x^2 - 6x + 1 = 0$, with roots $3 \pm 2\sqrt{2}$. Let $\rho = 3 + 2\sqrt{2} = 5.8284271247\ldots$. Then $c_{n+1} = \lfloor \rho c_n \rfloor$ for all $n \geq 0$,and $\lim_{n \to \infty} (c_{n+1}/c_n) = \rho$. To get the s_n corresponding to c_n (*not* the same values of n) we have $\left(\left\lfloor \frac{c_n}{\sqrt{2}} \right\rfloor\right)^2 + \left(\left\lceil \frac{c_n}{\sqrt{2}} \right\rceil\right)^2 = c_n^2$, where $\left\lfloor \frac{c_n}{\sqrt{2}} \right\rfloor + 1 = \left\lceil \frac{c_n}{\sqrt{2}} \right\rceil$. It is remarkable how close the (irrational!) values of ρ^n are to integers. (Actually, $\rho^n + \rho^{-n}$ *is* an integer for every $n \geq 1$, and the powers of $\rho^{-1} = 3 - 2\sqrt{2}$ go to 0 very rapidly with n.)

Reprinted from Vol. 55, No. 1, March 2005 issue of Information Theory Newsletter continued

6. If any s_n is a perfect *even* power it must be in the sequence C. Among the first 250 terms of $\{s_n\}$, the only power higher than the second power is $s_{119} = 28,561 = 13^4$. I don't know of any *odd* (perfect) powers in S, or other perfect even powers, but they may well exist.

7. It is "very likely" that the sequence $S = \{s_n\}$ contains infinitely many primes. However, this has not yet been proved for *any* quadratic expression in n. The "twin primes" result from removing, from the set of *all* odd integers > 0, two residue classes modulo every prime $p > 2$, and no one has yet been able to prove that there are infinitely many twin primes. In the quadratic sequence S, of a sparse subset of the odd integers > 0, we remove two residue classes modulo those primes of the form $4m + 1$ to see what (prime) values remain; so proving that S contains infinitely many prime values seems at least as hard as (or harder than) the "twin prime" problem.

Reprinted from Vol. 55, No. 2, June 2005 issue of Information Theory Newsletter



GOLOMB'S PUZZLE COLUMN™

Perfect Powers and Powerful Numbers

Solomon W. Golomb

The set P (the *perfect powers*) consists of all squares, cubes, and higher powers of the positive integers. The set Q (the *powerful numbers*) consists of those positive integers n for which, if a prime p divides n then p^2 divides n.

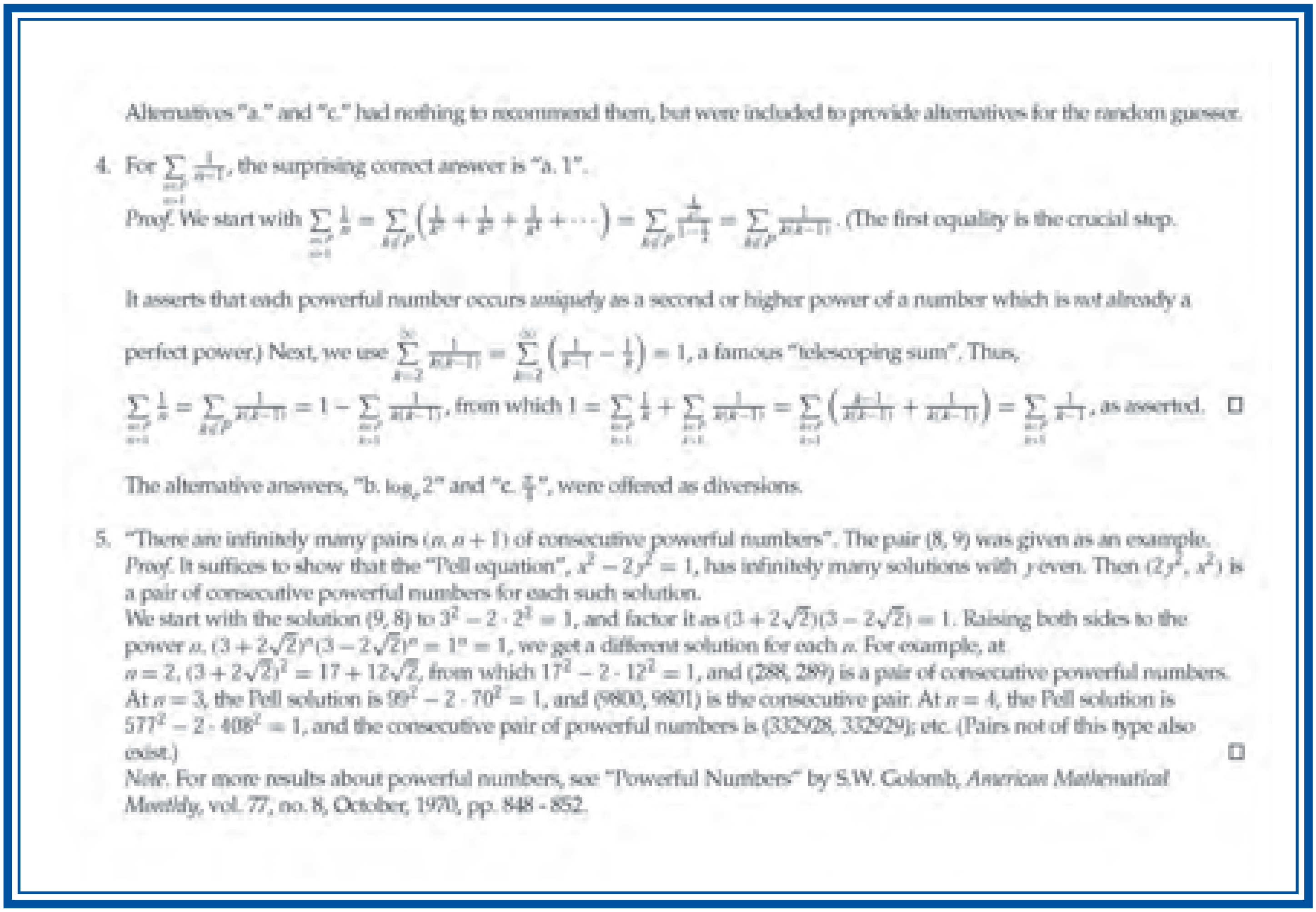
1. The relation between the sets P and Q is given by:
   c. Q = {set of all finite products of elements of P}.
   Proof. It is clear that every (finite) product of perfect powers is a powerful number. It is also clear that every powerful number is a finite product of powers (higher than the first power) of prime numbers, and therefore a finite product of elements of P. □
   The other alternatives offered do not work. For "a. $Q = P \times P$", the element $q = 2^2 \cdot 3^3 \cdot 5^5$ is in Q but not in $P \times P$. For "b. $Q = P + P$", $4 + 9 \in P + P$, but $13 \notin Q$.

2. The $\sum_{n \in Q} \frac{1}{n}$ is given by:

   c. $\zeta(2)\zeta(3)/\zeta(6)$.
   Proof. Every powerful number n can be written uniquely in the form $n = sc$, where s is a perfect square and c is a perfect cube, subject to the added condition that if p is a prime divisor of c, then p^3 is the exact power of p that divides c. The idea here is that with $n = \prod_{i=1}^{r} p_i^{a_i}$, where $p_1, p_2, \cdots, p_r$ are the distinct prime divisors of $n \in Q$, all the a_i's satisfy $a_i \geq 2$. When a_j is even, we use $p_j^{a_j}$ as a factor of s. When a_j is odd (and therefore ≥ 3), we use p_j^3 as a factor of c, and $p_j^{a_j-3}$ as a factor of s. (Since a_j is odd, $a_j - 3$ is even, so $p_j^{a_j-3}$ is a perfect square.) For example, if $n = 2^2 \cdot 3^3 \cdot 5^5$, we have $s = 2^2 \cdot 5^2 = 10^2$ and $c = 3^3 \cdot 5^3 = 15^3$, with $n = sc$.
   In view of this unique factorization of powerful numbers,
   $$\sum_{n \in Q} \frac{1}{n} = \left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right) \cdot \prod_{\text{all primes } p} \left(1 + \frac{1}{p^3}\right) = \zeta(2) \cdot (\zeta(3)/\zeta(6)).$$
   The relationship $\prod_{\text{all } p} \left(1 + \frac{1}{p^3}\right) = \zeta(3)/\zeta(6)$ follows from $\left(1 + \frac{1}{p^3}\right) = \left(1 - \frac{1}{p^6}\right) / \left(1 - \frac{1}{p^3}\right)$ and the Euler Product Formula whereby $\prod_{\text{all } p} \left(1 - \frac{1}{p^s}\right) = \frac{1}{\zeta(s)}$ for $s > 1$. □
   The other alternatives offered do not work. For "a. $\zeta(2) + \zeta(3) - \zeta(6)$", while the reciprocals of the squares and cubes are included, many other powerful numbers (e.g. $2^5, 3^5, 5^5, 2^7, \ldots$) do not have their reciprocals included. For "b. $\zeta(2)\zeta(3) - \zeta(6) + 1$", all the right reciprocals occur, but many occur more than once. For example, $\frac{1}{2^{36}}$ occurs as a product of a square reciprocal times a cube reciprocal in $\zeta(2)\zeta(3)$ in seven ways $(2^{36} = (2^{18})^2(1)^3 = (2^{15})^2(2^2)^3 = (2^{12})^2(2^4)^3 = (2^9)^2(2^6)^3 = (2^6)^2(2^8)^3 = (2^3)^2(2^{10})^3 = (1)^2(2^{12})^3)$, but is removed only once in $\zeta(6)$.

3. For $\sum_{n \in P} \frac{1}{n}$, the correct answer is "b. $-\sum_{k=2}^{\infty} \mu(k)\zeta(k)$."
   Proof. This is a typical "inclusion-exclusion" proof. The first term, $-\mu(2)\zeta(2) = \zeta(2) = \sum_{m=1}^{\infty} \frac{1}{m^2}$, gives the reciprocals of all the perfect squares. The next term, $-\mu(3)\zeta(3) = \zeta(3) = \sum_{m=1}^{\infty} \frac{1}{m^3}$, gives the reciprocals of all the perfect cubes; but then $-\mu(6)\zeta(6) = -\zeta(6) = -\sum_{m=1}^{\infty} \frac{1}{m^6}$ removes the reciprocals of the sixth powers, which appeared in both $\zeta(2)$ and $\zeta(3)$. Similarly, $-\mu(5)\zeta(5) = \zeta(5) = \sum_{m=1}^{\infty} \frac{1}{m^5}$ puts in the reciprocals of all fifth powers, but then $-\mu(10)\zeta(10) = -\sum_{m=1}^{\infty} \frac{1}{m^{10}}$ and $-\mu(15)\zeta(15) = -\sum_{m=1}^{\infty} \frac{1}{m^{15}}$ remove terms that were counted twice, and $-\mu(30)\zeta(30) = \sum_{m=1}^{\infty} \frac{1}{m^{30}}$ puts back terms once that appeared three times (in $\zeta(2)$, $\zeta(3)$, and $\zeta(15)$) but were also removed three times (in $-\zeta(6)$, $-\zeta(10)$, and $-\zeta(15)$), and so on. □

Reprinted from Vol. 55, No. 2, June 2005 issue of Information Theory Newsletter continued



Alternatives "a." and "c." had nothing to recommend them, but were included to provide alternatives for the random guesser.

4. For $\sum_{\substack{n\in P\\ n>1}} \frac{1}{n-1}$, the surprising correct answer is "a. 1".

   Proof. We start with $\sum_{\substack{n\in P\\ n>1}} \frac{1}{n} = \sum_{k\notin P}\left(\frac{1}{k^2}+\frac{1}{k^3}+\frac{1}{k^4}+\cdots\right) = \sum_{k\notin P}\frac{\frac{1}{k^2}}{1-\frac{1}{k}} = \sum_{k\notin P}\frac{1}{k(k-1)}$. (The first equality is the crucial step.

   It asserts that each powerful number occurs *uniquely* as a second or higher power of a number which is *not* already a perfect power.) Next, we use $\sum_{k=2}^{\infty}\frac{1}{k(k-1)} = \sum_{k=2}^{\infty}\left(\frac{1}{k-1}-\frac{1}{k}\right) = 1$, a famous "telescoping sum". Thus,

   $\sum_{\substack{n\in P\\ n>1}} \frac{1}{n} = \sum_{k\notin P}\frac{1}{k(k-1)} = 1 - \sum_{\substack{n\in P\\ k>1}}\frac{1}{k(k-1)}$, from which $1 = \sum_{\substack{k\in P\\ k>1}}\frac{1}{k} + \sum_{\substack{k\in P\\ k>1}}\frac{1}{k(k-1)} = \sum_{\substack{k\in P\\ k>1}}\left(\frac{k-1}{k(k-1)}+\frac{1}{k(k-1)}\right) = \sum_{\substack{n\in P\\ n>1}}\frac{1}{n-1}$, as asserted. □

   The alternative answers, "b. $\log_e 2$" and "c. $\frac{\pi}{4}$", were offered as diversions.

5. "There are infinitely many pairs $(n, n+1)$ of consecutive powerful numbers". The pair (8, 9) was given as an example.
   Proof. It suffices to show that the "Pell equation", $x^2 - 2y^2 = 1$, has infinitely many solutions with y even. Then $(2y^2, x^2)$ is a pair of consecutive powerful numbers for each such solution.
   We start with the solution (9, 8) to $3^2 - 2\cdot 2^2 = 1$, and factor it as $(3+2\sqrt{2})(3-2\sqrt{2}) = 1$. Raising both sides to the power n, $(3+2\sqrt{2})^n(3-2\sqrt{2})^n = 1^n = 1$, we get a different solution for each n. For example, at $n = 2$, $(3+2\sqrt{2})^2 = 17 + 12\sqrt{2}$, from which $17^2 - 2\cdot 12^2 = 1$, and (288, 289) is a pair of consecutive powerful numbers. At $n = 3$, the Pell solution is $99^2 - 2\cdot 70^2 = 1$, and (9800, 9801) is the consecutive pair. At $n = 4$, the Pell solution is $577^2 - 2\cdot 408^2 = 1$, and the consecutive pair of powerful numbers is (332928, 332929); etc. (Pairs not of this type also exist.) □
   Note. For more results about powerful numbers, see "Powerful Numbers" by S.W. Golomb, *American Mathematical Monthly*, vol. 77, no. 8, October, 1970, pp. 848 - 852.

Reprinted from Vol. 55, No. 3, September 2005 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Some Matrix Questions Solutions

Solomon W. Golomb

1. The matrices $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ are similar, since $P^{-1}AP = B$ with $P = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ and $P^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. However, $AB = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$ and $BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ are not similar, since clearly $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ is similar only to itself.

2. *Theorem*. If, for a given complex matrix M, there exists a unitary matrix U such that $U^{-1}MU = \Lambda$, where Λ is a diagonal matrix, then M is normal.

   Proof. From $U^{-1}MU = \Lambda$, we have $\Lambda^* = \Lambda^H = (U^{-1}MU)^H = U^H M^H (U^{-1})^H = U^{-1}M^H U$. Now $\Lambda\Lambda^* = \Lambda^*\Lambda$, because if Λ is the diagonal matrix with $\lambda_1, \lambda_2, \ldots, \lambda_n$ as its diagonal elements, then Λ^* is the diagonal matrix with $\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*$ as its diagonal elements, and both $\Lambda\Lambda^*$ and $\Lambda^*\Lambda$ are diagonal matrices with $|\lambda_1|^2, |\lambda_2|^2, \ldots, |\lambda_n|^2$ as their diagonal elements. Now $\Lambda\Lambda^* = (U^{-1}MU)(U^{-1}M^H U) = U^{-1}(MM^H)U$, $\Lambda^*\Lambda = (U^{-1}M^H U)(U^{-1}MU) = U^{-1}(M^H M)U$, and since these are equal, $MM^H = M^H M$, so M is normal. □

3. "If N_1 and N_2 are normal $n \times n$ matrices then $N_1 N_2$ is normal" is *false*. For a counter-example, we can use the fact that every (real) symmetric matrix S is normal, since $S^H = S^T = S$, from which $SS^H = S^2 = S^H S$. Then with $N_1 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$ and $N_2 = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$, we have $P = N_1 N_2 = \begin{pmatrix} 4 & 9 \\ 7 & 14 \end{pmatrix}$ and $P^H = P^T = N_2^T N_1^T = N_2 N_1 = \begin{pmatrix} 4 & 7 \\ 9 & 14 \end{pmatrix}$. Now $PP^H = \begin{pmatrix} 4 & 9 \\ 7 & 14 \end{pmatrix} \times \begin{pmatrix} 4 & 7 \\ 9 & 14 \end{pmatrix} = \begin{pmatrix} 97 & 154 \\ 154 & 245 \end{pmatrix}$ but $P^H P = \begin{pmatrix} 4 & 7 \\ 9 & 14 \end{pmatrix}\begin{pmatrix} 4 & 9 \\ 7 & 14 \end{pmatrix} = \begin{pmatrix} 65 & 134 \\ 134 & 277 \end{pmatrix}$, so that $P = N_1 N_2$ is not normal.

4. Let $R = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \right\} = \{O, Z, I, J\}$ over $GF(2)$. R forms a commutative group with respect to matrix addition modulo 2, and R is closed under matrix multiplication modulo 2. In this ring, both I and J are "left identities", since $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$, but neither one is a "right identity", since $ZI = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$, and $ZJ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = O$.

5. "If the n^2 elements of an $n \times n$ matrix A are integers chosen independently and at random, what is the probability that $|A|$, the determinant of A, is odd?"

   For all $n \geq 2$, the answer is *not* one-half. Rather than worry about what sample space integers can be chosen from "independently and at random", we need only agree that each entry in A is equally likely to be even or odd, and independently of the other entries. Then our question is equivalent to: "What fraction of the $n \times n$ matrices over $GF(2)$ are non-singular?" (The *even* determinants all reduce to 0, and the *odd* determinants all reduce to 1, modulo 2.) To form a non-singular $n \times n$ matrix over $GF(2)$, we can pick the top row in $2^n - 1$ ways (only the all-zeroes case is excluded), the second row in $2^n - 2$ ways, and the j^{th} row in $2^n - 2^{j-1}$ ways, for all j, $1 \leq j \leq n$. This gives $\prod_{j=1}^{n}(2^n - 2^{j-1})$ non-singular matrices, out of $2^{(n^2)}$ matrices altogether. Then the probability is the ratio, which simplifies to $\prod_{j=1}^{n}(1 - 2^{-j}) = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{7}{8} \cdots \frac{2^n - 1}{2^n}$. This sequence of probabilities, $\frac{1}{2}, \frac{3}{8}, \frac{21}{64}, \frac{315}{1024}, \cdots$, converges rather rapidly to a positive limiting value, $\prod_{j=1}^{\infty}\left(1 - \frac{1}{2^j}\right) = 0.2887878\ldots$. Thus, the probability that a "large" $n \times n$ matrix of "random" integers would have an odd determinant is a bit less than 29%. (At $n = 8$, this probability is already down to $0.289919\ldots$.)

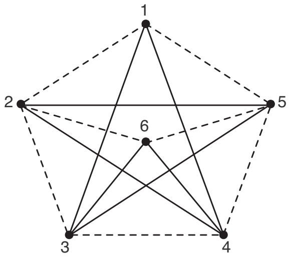
Reprinted from Vol. 55, No. 4, December 2005 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Some Matrix Questions Solutions

Solomon W. Golomb

1. It is sufficient to remove only one edge from K_6 so that the fourteen remaining edges can be colored using two colors without forming a solid-color triangle:



Only the edge connecting the points **1** and **6** is missing from K_6. The clever way to view this example is to consider that we started with a triangle-free 2-coloring of K_5, on the points **1**, **2**, **3**, **4**, and **5**, and then adjoined **6** as a "clone" of point **1**. That is, **6** is connected to each of **2**, **3**, **4**, and **5** with the same colors of edges as those emanating from **1**; so if a solid-color triangle involving **6** is formed, there would already have been a solid-color triangle involving **1**.

2. In a similar way, we can adjoin clones for *each* of the five original points **1**, **2**, **3**, **4**, **5**. Call these new points **1'**, **2'**, **3'**, **4'**, **5'**. Then the only lines missing from K_{10} are the five lines connecting each original point to its clone. That is, 40 of the 45 edges of K_{10} can be 2-colored without forming a solid-color triangle. (To convince yourself that no forbidden triangles are formed, adjoin the new "clone points" one at a time.)

3. The question told you that all the edges of K_{16} can be 3-colored without forming a solid-color triangle, but that this is not true for K_{17}. Hence, we can adjoin a clone P' for one of the points P of K_{16}, and connect P' to every original point Q of K_{16} except P, with the same color edge as the edge connecting P to Q, without forming a solid-color triangle. That is, we need remove only one edge from K_{17} so that the remaining 135 edges can be 3-colored without forming a solid-color triangle.

4. The same reasoning shows that if $r = r(c)$ is the *smallest* positive integer such that, if the $\binom{r}{2}$ edges of K_r are colored using c colors a solid-color triangle must be created, then it suffices to remove a single edge from K_r so that the remaining edges can be c-colored without forming a solid-color triangle. Specifically, we start with a c-coloring of K_{r-1} that has no solid-color triangle, and adjoin a new point P' as a clone of P, one of the original $r-1$ points of K_{r-1}, and proceed as in Solutions 1 and 3 above.

Note. A key idea for this column was supplied by Herbert Taylor.

Reprinted from Vol. 56, No. 1, March 2006 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Simple Probabilities Solution

Solomon W. Golomb

1. After the first four cards dealt are all seen to be hearts, the deck still contains 9 hearts among 48 cards, so the probability that the fifth card will also be a heart is $\frac{9}{48} = \frac{3}{16} = 0.1875$.

2. The number of ways to select 5 cards, all hearts, is $\binom{13}{5} = 1287$. The number of 5-card hands with *at least* four hearts is the number of hands with 5 hearts *plus* the number of hands with 4 hearts and one non-heart, which equals $\binom{13}{5} + \binom{13}{4}\binom{39}{1} = 1287 + 27,885 = 29,172$. Thus the probability that a 5-card hand with at least four hearts actually contains five hearts is $\frac{1287}{29,172} = \frac{3}{68} = 0.044117647\ldots$.

3. (a) For *at least* one of six dice to show a **5**, the probability is $1 - \left(\frac{5}{6}\right)^6 = \frac{31,031}{46,656} = 0.6651020233\ldots$, or nearly two-thirds.

   (b) For *exactly* one of six dice to show a **5**, there is a choice of *which* of the six dice shows the **5**, and the other dice must *avoid* showing a **5**; so the probability is: $\frac{6\times 5^5}{6^6} = \frac{5^5}{6^5} = \frac{3125}{7776} = 0.40187757\ldots$.

4. (a) The probability that all six dice turn up the same is $\frac{6}{6^6} = \frac{1}{7776} = 0.000128601\ldots$.

   (b) The probability that all six dice turn up different (i.e. that every number from 1 to 6 appears) is $\frac{6!}{6^6} = \frac{5}{324} = 0.015432099\ldots$.

5. Originally, your chance of guessing right is one in four, or 25%. After two wrong alternatives are removed, your original choice is still only 25% right, so by switching you increase your winning probability to 75%. (The two remaining doors are *not* equally likely!)

6. Among the four honest coins plus one two-headed coin, there are six head faces, two of which belong to the crooked coin; so the probability that the unseen side is also heads is $\frac{2}{6} = 0.333333\ldots$.

7. Expected number of tosses of a pair of dice to see a total of k, $2 \le k \le 12$, on the two dice, is $E(k)$, where

k	$E(k)$	k	$E(k)$	k	$E(k)$
2	24.6051	5	5.8849	9	5.8849
3	12.1268	6	4.6355	10	7.9662
4	7.9662	7	3.8018	11	12.1268
		8	4.6355	12	24.6051

   $E(k)$ is the solution to $\left(\frac{36-k+1}{36}\right)^{E(k)} = \frac{1}{2}$, and is given by $E(k) = \log 2/(\log 36 - \log(36-k+1))$ for $2 \le k \le 7$, and $E(7+a) = E(7-a)$ for $1 \le a \le 5$.

8. Although the grass always looks greener on the other side of the fence, it can't be true mathematically that *no matter which* of the two envelopes you picked, the *other* one is 25% better (at least in expectation)! So where is the fallacy? The problem stated that x, a positive real number, was picked "at random", but it didn't specify the *distribution* from which x was selected. The reasoning tacitly assumed it was from the "uniform distribution", but there is no uniform distribution on the positive real numbers. Your best strategy is to assume (best guess) what the *mean* of the unrevealed distribution is, accept y if it exceeds this mean, but switch if it is below. Another defensible strategy is to accept y if you would be satisfied with that amount of money, but reject y otherwise (the "minimum regret" approach).

Reprinted from Vol. 56, No. 2, June 2006 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Mini-Sudoku Solution

Solomon W. Golomb

1. There are 288 = 12 × 4! distinct Mini-Sudoku solutions. The factor 4! = 24 corresponds to all permutations of the four symbols. Here are the 12 cases that differ by more than permutation of the symbols.

1.

1	2	3	4
3	4	1	2
2	1	4	3
4	3	2	1

2.

1	2	3	4
3	4	1	2
2	3	4	1
4	1	2	3

3.

1	2	3	4
3	4	1	2
4	1	2	3
2	3	4	1

4.

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

5.

1	2	3	4
3	4	2	1
2	1	4	3
4	3	1	2

6.

1	2	3	4
3	4	2	1
4	3	1	2
2	1	4	3

7.

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

8.

1	2	3	4
4	3	2	1
2	4	1	3
3	1	4	2

9.

1	2	3	4
4	3	2	1
3	1	4	2
2	4	1	3

10.

1	2	3	4
4	3	2	1
3	4	1	2
2	1	4	3

11.

1	2	3	4
4	3	1	2
2	1	4	3
3	4	2	1

12.

1	2	3	4
4	3	1	2
3	4	2	1
2	1	4	3

2. Here is a Mini-Sudoku solution in which the four elements on each of the two diagonals are also distinct: (It is the seventh of the twelve cases shown above. The fourth case also has this property.)

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

3. At least four cells must be filled in to guarantee a unique Mini-Sudoku solution. Three distinct symbols must appear, since otherwise two unused symbols could be interchanged in the filled-in solution, destroying uniqueness. I tried all inequivalent ways of placing one each of 1, 2, and 3 in the 4 × 4 grid, and none of these led to a unique Mini-Sudoku solution. There are many ways to place four symbols that will guarantee a unique solution. Here is one of them:

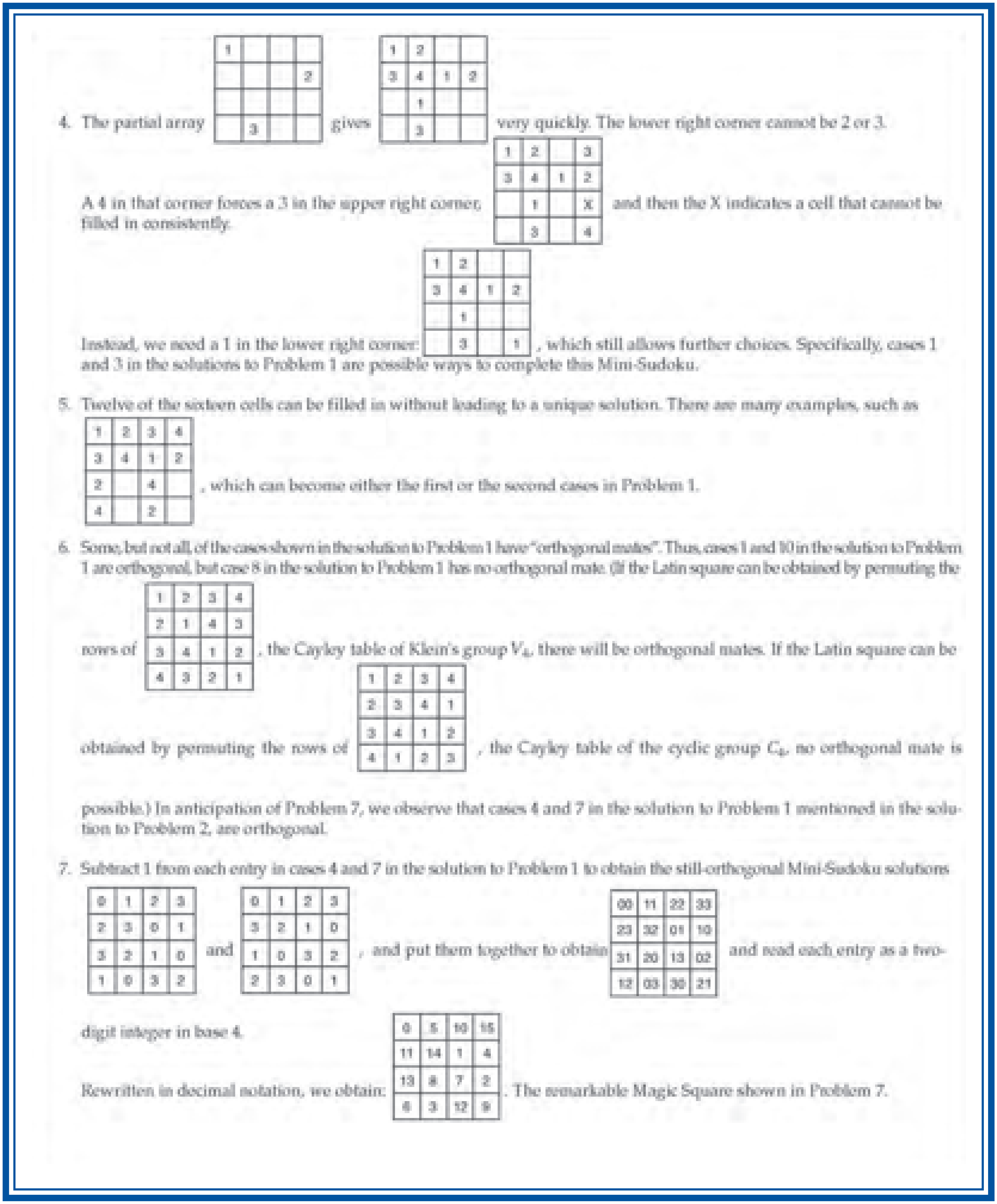
1			
			2
	3		
		3	

, which forces

1	2	3	4
3	4	1	2
4	3	2	1
2	1	3	4

.

Reprinted from Vol. 56, No. 2, June 2006 issue of Information Theory Newsletter continued



4. The partial array gives very quickly. The lower right corner cannot be 2 or 3.

1			
			2
	3		

1	2		
3	4	1	2
	1		
	3		

A 4 in that corner forces a 3 in the upper right corner, and then the X indicates a cell that cannot be filled in consistently.

1	2		3
3	4	1	2
	1		X
	3		4

Instead, we need a 1 in the lower right corner: , which still allows further choices. Specifically, cases 1 and 3 in the solutions to Problem 1 are possible ways to complete this Mini-Sudoku.

1	2		
3	4	1	2
	1		
	3		1

5. Twelve of the sixteen cells can be filled in without leading to a unique solution. There are many examples, such as , which can become either the first or the second cases in Problem 1.

1	2	3	4
3	4	1	2
2		4	
4		2	

6. Some, but not all, of the cases shown in the solution to Problem 1 have "orthogonal mates". Thus, cases 1 and 10 in the solution to Problem 1 are orthogonal, but case 8 in the solution to Problem 1 has no orthogonal mate. (If the Latin square can be obtained by permuting the rows of , the Cayley table of Klein's group V_4, there will be orthogonal mates. If the Latin square can be obtained by permuting the rows of , the Cayley table of the cyclic group C_4, no orthogonal mate is possible.) In anticipation of Problem 7, we observe that cases 4 and 7 in the solution to Problem 1 mentioned in the solution to Problem 2, are orthogonal.

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

7. Subtract 1 from each entry in cases 4 and 7 in the solution to Problem 1 to obtain the still-orthogonal Mini-Sudoku solutions and , and put them together to obtain and read each entry as a two-digit integer in base 4.

0	1	2	3
2	3	0	1
3	2	1	0
1	0	3	2

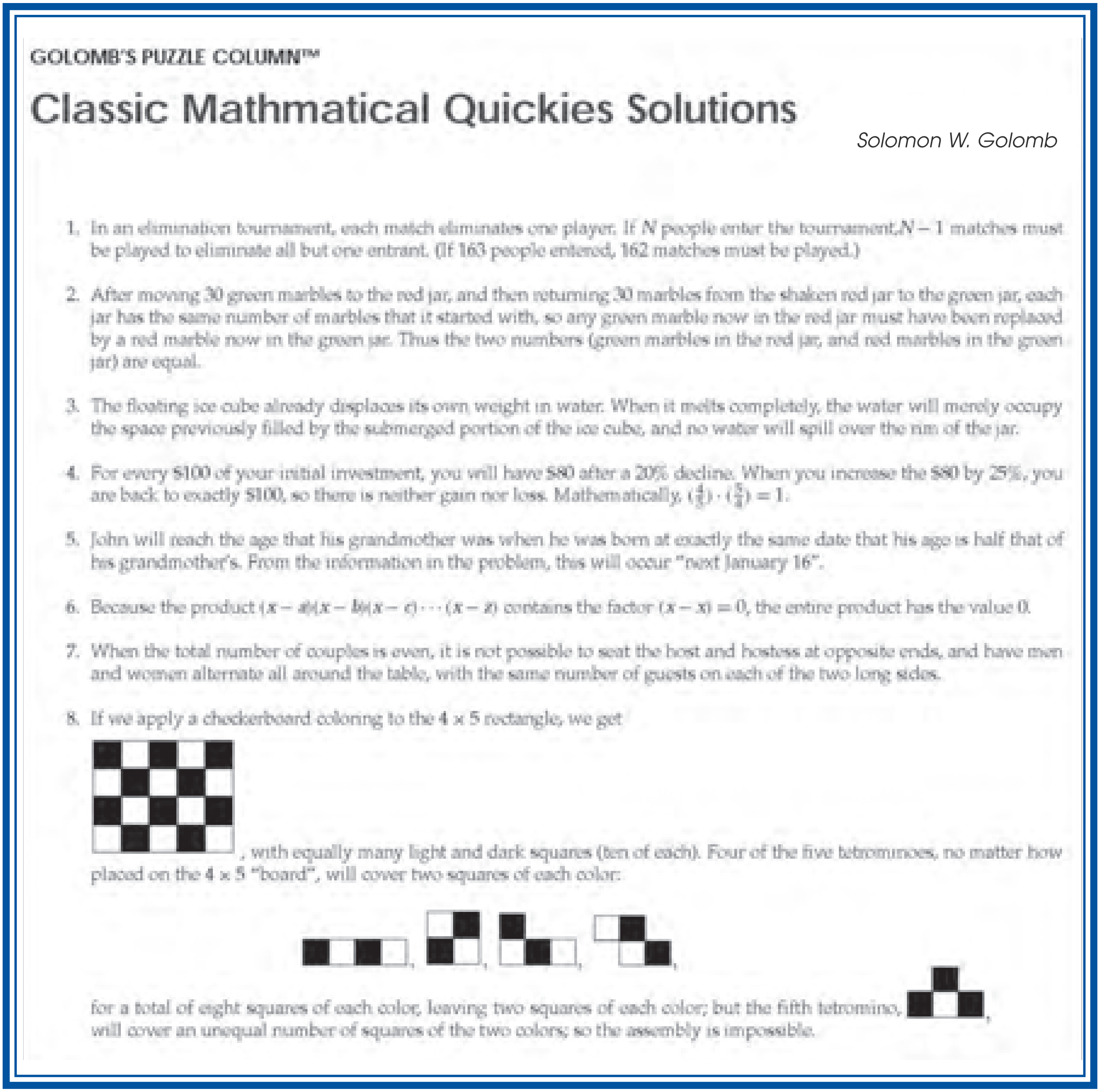
0	1	2	3
3	2	1	0
1	0	3	2
2	3	0	1

00	11	22	33
23	32	01	10
31	20	13	02
12	03	30	21

Rewritten in decimal notation, we obtain: . The remarkable Magic Square shown in Problem 7.

0	5	10	15
11	14	1	4
13	8	7	2
6	3	12	9

Reprinted from Vol. 56, No. 3, September 2006 issue of Information Theory Newsletter



GOLOMB'S PUZZLE COLUMN™

Classic Mathmatical Quickies Solutions

Solomon W. Golomb

1. In an elimination tournament, each match eliminates one player. If N people enter the tournament, $N-1$ matches must be played to eliminate all but one entrant. (If 163 people entered, 162 matches must be played.)

2. After moving 30 green marbles to the red jar, and then returning 30 marbles from the shaken red jar to the green jar, each jar has the same number of marbles that it started with, so any green marble now in the red jar must have been replaced by a red marble now in the green jar. Thus the two numbers (green marbles in the red jar, and red marbles in the green jar) are equal.

3. The floating ice cube already displaces its own weight in water. When it melts completely, the water will merely occupy the space previously filled by the submerged portion of the ice cube, and no water will spill over the rim of the jar.

4. For every $100 of your initial investment, you will have $80 after a 20% decline. When you increase the $80 by 25%, you are back to exactly $100, so there is neither gain nor loss. Mathematically, $(\frac{4}{5}) \cdot (\frac{5}{4}) = 1$.

5. John will reach the age that his grandmother was when he was born at exactly the same date that his age is half that of his grandmother's. From the information in the problem, this will occur "next January 16".

6. Because the product $(x-a)(x-b)(x-c) \cdots (x-z)$ contains the factor $(x-x) = 0$, the entire product has the value 0.

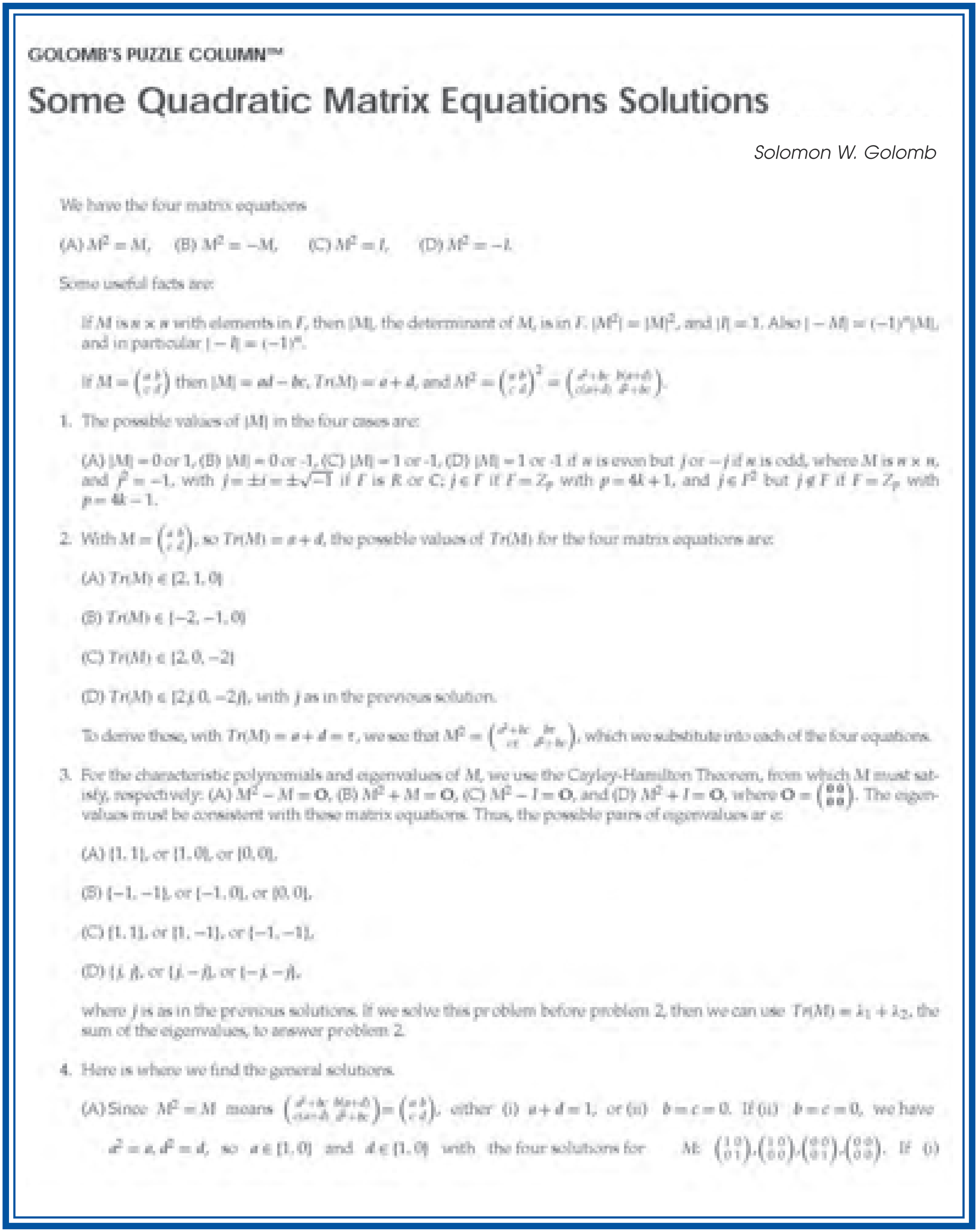
7. When the total number of couples is even, it is not possible to seat the host and hostess at opposite ends, and have men and women alternate all around the table, with the same number of guests on each of the two long sides.

8. If we apply a checkerboard coloring to the 4 × 5 rectangle, we get , with equally many light and dark squares (ten of each). Four of the five tetrominoes, no matter how placed on the 4 × 5 "board", will cover two squares of each color:

, , , ,

for a total of eight squares of each color, leaving two squares of each color; but the fifth tetromino, , will cover an unequal number of squares of the two colors; so the assembly is impossible.

Reprinted from Vol. 56, No. 4, December 2006 issue of Information Theory Newsletter



GOLOMB'S PUZZLE COLUMN™

Some Quadratic Matrix Equations Solutions

Solomon W. Golomb

We have the four matrix equations

(A) $M^2 = M$, (B) $M^2 = -M$, (C) $M^2 = I$, (D) $M^2 = -I$.

Some useful facts are:

If M is $n \times n$ with elements in F, then $|M|$, the determinant of M, is in F. $|M^2| = |M|^2$, and $|I| = 1$. Also $|-M| = (-1)^n|M|$, and in particular $|-I| = (-1)^n$.

If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $|M| = ad - bc$, $Tr(M) = a + d$, and $M^2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & d^2+bc \end{pmatrix}$.

1. The possible values of $|M|$ in the four cases are:

   (A) $|M| = 0$ or 1, (B) $|M| = 0$ or -1, (C) $|M| = 1$ or -1, (D) $|M| = 1$ or -1 if n is even but j or $-j$ if n is odd, where M is $n \times n$, and $j^2 = -1$, with $j = \pm i = \pm\sqrt{-1}$ if F is R or C; $j \in F$ if $F = Z_p$ with $p = 4k+1$, and $j \in F^2$ but $j \notin F$ if $F = Z_p$ with $p = 4k-1$.

2. With $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so $Tr(M) = a + d$, the possible values of $Tr(M)$ for the four matrix equations are:

   (A) $Tr(M) \in \{2, 1, 0\}$

   (B) $Tr(M) \in \{-2, -1, 0\}$

   (C) $Tr(M) \in \{2, 0, -2\}$

   (D) $Tr(M) \in \{2j, 0, -2j\}$, with j as in the previous solution.

   To derive these, with $Tr(M) = a + d = \tau$, we see that $M^2 = \begin{pmatrix} a^2+bc & b\tau \\ c\tau & d^2+bc \end{pmatrix}$, which we substitute into each of the four equations.

3. For the characteristic polynomials and eigenvalues of M, we use the Cayley-Hamilton Theorem, from which M must satisfy, respectively: (A) $M^2 - M = \mathbf{O}$, (B) $M^2 + M = \mathbf{O}$, (C) $M^2 - I = \mathbf{O}$, and (D) $M^2 + I = \mathbf{O}$, where $\mathbf{O} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. The eigenvalues must be consistent with these matrix equations. Thus, the possible pairs of eigenvalues ar e:

   (A) $\{1, 1\}$, or $\{1, 0\}$, or $\{0, 0\}$,

   (B) $\{-1, -1\}$, or $\{-1, 0\}$, or $\{0, 0\}$,

   (C) $\{1, 1\}$, or $\{1, -1\}$, or $\{-1, -1\}$,

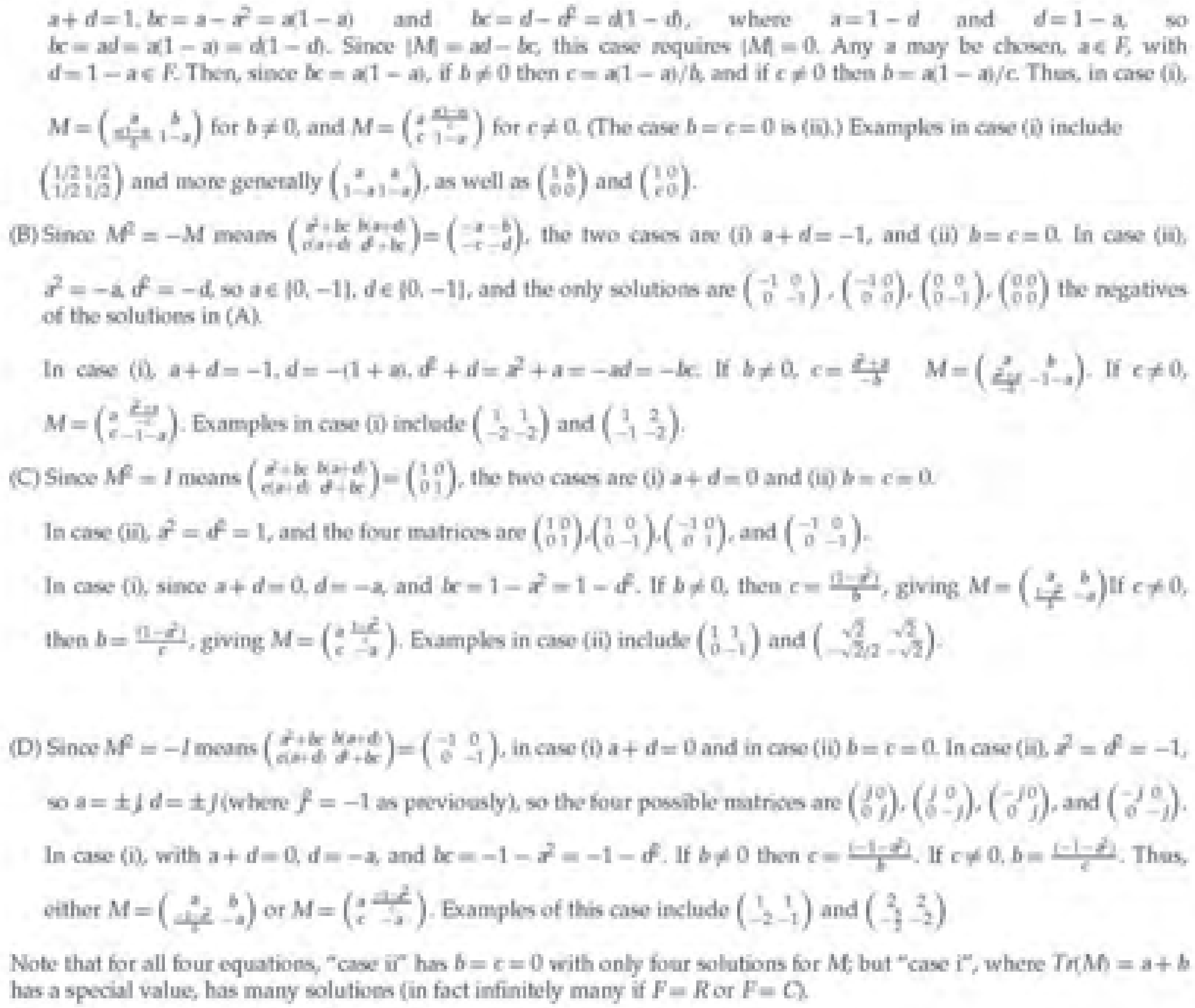
   (D) $\{j, j\}$, or $\{j, -j\}$, or $\{-j, -j\}$,

   where j is as in the previous solutions. If we solve this pr oblem before problem 2, then we can use $Tr(M) = \lambda_1 + \lambda_2$, the sum of the eigenvalues, to answer problem 2.

4. Here is where we find the general solutions.

   (A) Since $M^2 = M$ means $\begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & d^2+bc \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, either (i) $a + d = 1$, or (ii) $b = c = 0$. If (ii) $b = c = 0$, we have $a^2 = a, d^2 = d$, so $a \in \{1, 0\}$ and $d \in \{1, 0\}$ with the four solutions for M: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. If (i)

Reprinted from Vol. 56, No. 4, December 2006 issue of Information Theory Newsletter continued



$a+d=1$, $bc=a-a^2=a(1-a)$ and $bc=d-d^2=d(1-d)$, where $a=1-d$ and $d=1-a$, so $bc=ad=a(1-a)=d(1-d)$. Since $|M|=ad-bc$, this case requires $|M|=0$. Any a may be chosen, $a\in F$, with $d=1-a\in F$. Then, since $bc=a(1-a)$, if $b\neq 0$ then $c=a(1-a)/b$, and if $c\neq 0$ then $b=a(1-a)/c$. Thus, in case (i),

$M=\begin{pmatrix} a & b \\ \frac{a(1-a)}{b} & 1-a \end{pmatrix}$ for $b\neq 0$, and $M=\begin{pmatrix} a & \frac{a(1-a)}{c} \\ c & 1-a \end{pmatrix}$ for $c\neq 0$. (The case $b=c=0$ is (ii).) Examples in case (i) include

$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and more generally $\begin{pmatrix} a & a \\ 1-a & 1-a \end{pmatrix}$, as well as $\begin{pmatrix} 1 & b \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ c & 0 \end{pmatrix}$.

(B) Since $M^2=-M$ means $\begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & d^2+bc \end{pmatrix}=\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$, the two cases are (i) $a+d=-1$, and (ii) $b=c=0$. In case (ii),

$a^2=-a$, $d^2=-d$, so $a\in\{0,-1\}$, $d\in\{0,-1\}$, and the only solutions are $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ the negatives of the solutions in (A).

In case (i), $a+d=-1$, $d=-(1+a)$, $d^2+d=a^2+a=-ad=-bc$. If $b\neq 0$, $c=\frac{a^2+a}{-b}$ $M=\begin{pmatrix} a & b \\ \frac{a^2+a}{-b} & -1-a \end{pmatrix}$. If $c\neq 0$,

$M=\begin{pmatrix} a & \frac{a^2+a}{-c} \\ c & -1-a \end{pmatrix}$. Examples in case (i) include $\begin{pmatrix} 1 & 1 \\ -2 & -2 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 \\ -1 & -2 \end{pmatrix}$.

(C) Since $M^2=I$ means $\begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & d^2+bc \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the two cases are (i) $a+d=0$ and (ii) $b=c=0$.

In case (ii), $a^2=d^2=1$, and the four matrices are $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, and $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$.

In case (i), since $a+d=0$, $d=-a$, and $bc=1-a^2=1-d^2$. If $b\neq 0$, then $c=\frac{(1-a^2)}{b}$, giving $M=\begin{pmatrix} a & b \\ \frac{1-a^2}{b} & -a \end{pmatrix}$If $c\neq 0$,

then $b=\frac{(1-a^2)}{c}$, giving $M=\begin{pmatrix} a & \frac{1-a^2}{c} \\ c & -a \end{pmatrix}$. Examples in case (ii) include $\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$ and $\begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2}/2 & -\sqrt{2} \end{pmatrix}$.

(D) Since $M^2=-I$ means $\begin{pmatrix} a^2+bc & b(a+d) \\ c(a+d) & d^2+bc \end{pmatrix}=\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, in case (i) $a+d=0$ and in case (ii) $b=c=0$. In case (ii), $a^2=d^2=-1$,

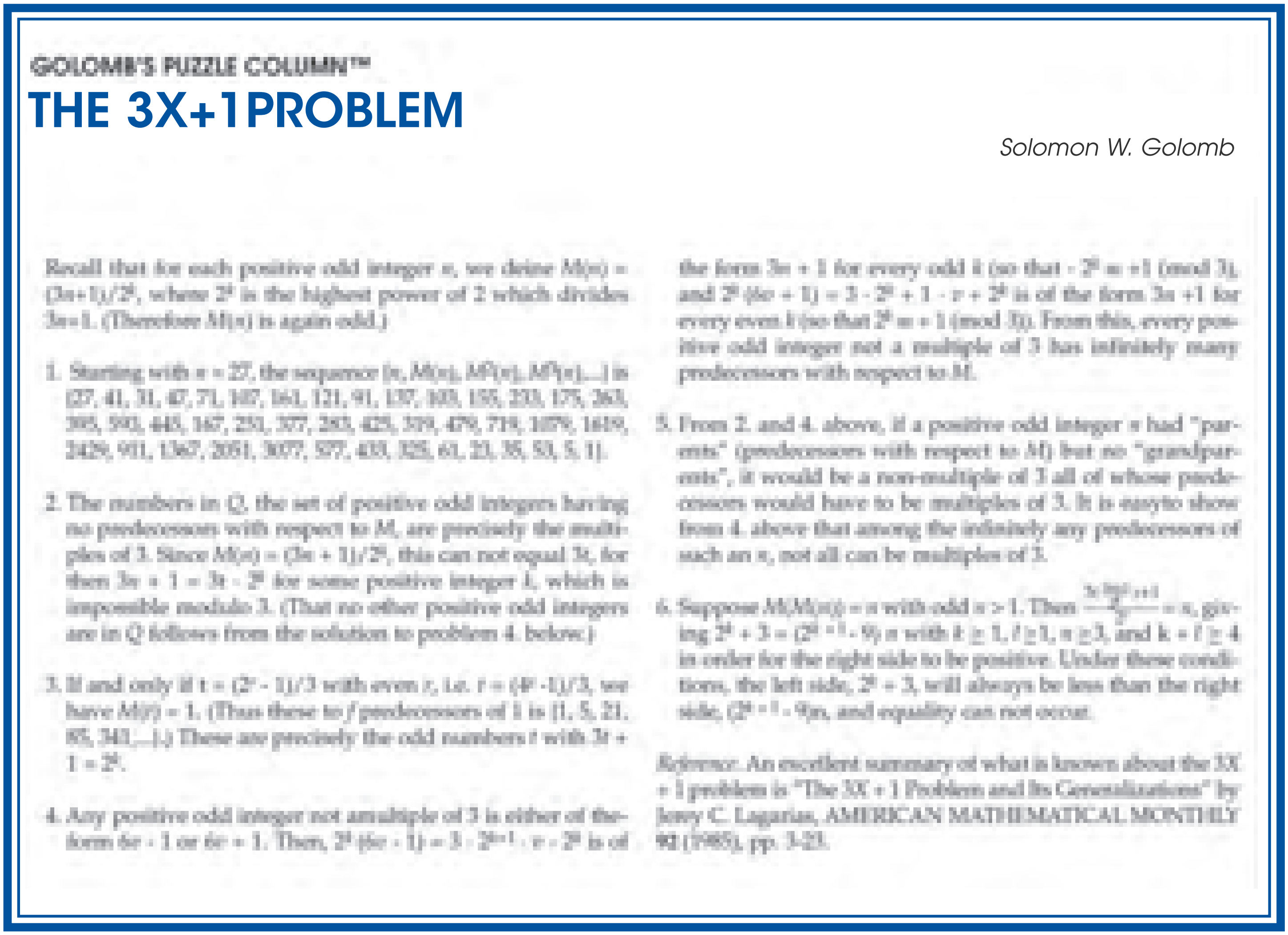
so $a=\pm j$, $d=\pm j$ (where $j^2=-1$ as previously), so the four possible matrices are $\begin{pmatrix} j & 0 \\ 0 & j \end{pmatrix}$, $\begin{pmatrix} j & 0 \\ 0 & -j \end{pmatrix}$, $\begin{pmatrix} -j & 0 \\ 0 & j \end{pmatrix}$, and $\begin{pmatrix} -j & 0 \\ 0 & -j \end{pmatrix}$.

In case (i), with $a+d=0$, $d=-a$, and $bc=-1-a^2=-1-d^2$. If $b\neq 0$ then $c=\frac{(-1-a^2)}{b}$. If $c\neq 0$, $b=\frac{(-1-a^2)}{c}$. Thus,

either $M=\begin{pmatrix} a & b \\ \frac{-1-a^2}{b} & -a \end{pmatrix}$ or $M=\begin{pmatrix} a & \frac{-1-a^2}{c} \\ c & -a \end{pmatrix}$. Examples of this case include $\begin{pmatrix} 1 & 1 \\ -2 & -1 \end{pmatrix}$ and $\begin{pmatrix} 2 & 2 \\ -\frac{5}{2} & -2 \end{pmatrix}$

Note that for all four equations, "case ii" has $b=c=0$ with only four solutions for M; but "case i", where $Tr(M)=a+b$ has a special value, has many solutions (in fact infinitely many if $F=R$ or $F=C$).

Reprinted from Vol. 57, No. 1, March 2007 issue of Information Theory Newsletter



GOLOMB'S PUZZLE COLUMN™

THE 3X+1PROBLEM

Solomon W. Golomb

Recall that for each positive odd integer n, we define $M(n) = (3n+1)/2^k$, where 2^k is the highest power of 2 which divides $3n+1$. (Therefore $M(n)$ is again odd.)

1. Starting with $n = 27$, the sequence $\{n, M(n), M^2(n), M^3(n),...\}$ is $\{27, 41, 31, 47, 71, 107, 161, 121, 91, 137, 103, 155, 233, 175, 263, 395, 593, 445, 167, 251, 377, 283, 425, 319, 479, 719, 1079, 1619, 2429, 911, 1367, 2051, 3077, 577, 433, 325, 61, 23, 35, 53, 5, 1\}$.

2. The numbers in Q, the set of positive odd integers having no predecessors with respect to M, are precisely the multiples of 3. Since $M(n) = (3n + 1)/2^k$, this can not equal $3t$, for then $3n + 1 = 3t \cdot 2^k$ for some positive integer k, which is impossible modulo 3. (That no other positive odd integers are in Q follows from the solution to problem 4. below.)

3. If and only if $t = (2^r - 1)/3$ with even r, i.e. $t = (4^s - 1)/3$, we have $M(t) = 1$. (Thus these to f predecessors of 1 is $\{1, 5, 21, 85, 341,...\}$.) These are precisely the odd numbers t with $3t + 1 = 2^k$.

4. Any positive odd integer not amultiple of 3 is either of the- form $6r - 1$ or $6r + 1$. Then, $2^k (6r - 1) = 3 \cdot 2^{k+1} \cdot r - 2^k$ is of the form $3n + 1$ for every odd k (so that $- 2^k \equiv +1 \pmod 3$), and $2^k (6r + 1) = 3 \cdot 2^{k+1} \cdot r + 2^k$ is of the form $3n +1$ for every even k (so that $2^k \equiv + 1 \pmod 3$). From this, every positive odd integer not a multiple of 3 has infinitely many predecessors with respect to M.

5. From 2. and 4. above, if a positive odd integer n had "parents" (predecessors with respect to M) but no "grandparents", it would be a non-multiple of 3 all of whose predecessors would have to be multiples of 3. It is easyto show from 4. above that among the infinitely any predecessors of such an n, not all can be multiples of 3.

6. Suppose $M(M(n)) = n$ with odd $n > 1$. Then $\frac{3(\frac{3n+1}{2^k})+1}{2^l} = n$, giving $2^k + 3 = (2^{k+l} - 9) n$ with $k \geq 1$, $l \geq 1$, $n \geq 3$, and $k + l \geq 4$ in order for the right side to be positive. Under these conditions, the left side, $2^k + 3$, will always be less than the right side, $(2^{k+l} - 9)n$, and equality can not occur.

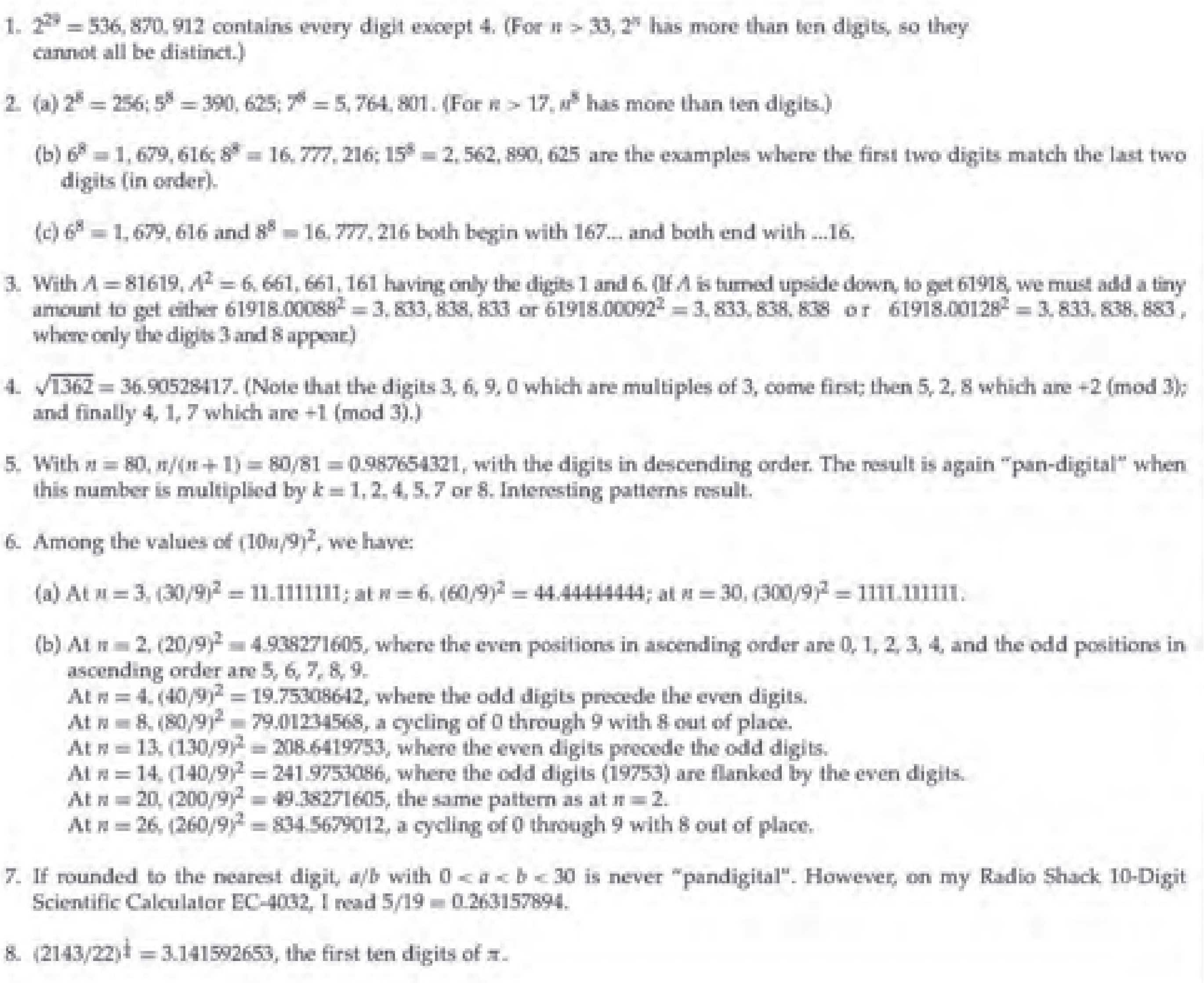
Reference. An excellent summary of what is known about the 3X + 1 problem is "The 3X + 1 Problem and Its Generalizations" by Jerry C. Lagarias, AMERICAN MATHEMATICAL MONTHLY 92 (1985), pp. 3-23.

Reprinted from Vol. 57, No. 2, June 2007 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

Calculator Magic Solutions

Solomon W. Golomb



1. $2^{29} = 536,870,912$ contains every digit except 4. (For $n > 33$, 2^n has more than ten digits, so they cannot all be distinct.)

2. (a) $2^8 = 256$; $5^8 = 390,625$; $7^8 = 5,764,801$. (For $n > 17$, n^8 has more than ten digits.)

   (b) $6^8 = 1,679,616$; $8^8 = 16,777,216$; $15^8 = 2,562,890,625$ are the examples where the first two digits match the last two digits (in order).

   (c) $6^8 = 1,679,616$ and $8^8 = 16,777,216$ both begin with 167... and both end with ...16.

3. With $A = 81619$, $A^2 = 6,661,661,161$ having only the digits 1 and 6. (If A is turned upside down, to get 61918, we must add a tiny amount to get either $61918.00088^2 = 3,833,838,833$ or $61918.00092^2 = 3,833,838,838$ o r $61918.00128^2 = 3,833,838,883$, where only the digits 3 and 8 appear.)

4. $\sqrt{1362} = 36.90528417$. (Note that the digits 3, 6, 9, 0 which are multiples of 3, come first; then 5, 2, 8 which are +2 (mod 3); and finally 4, 1, 7 which are +1 (mod 3).)

5. With $n = 80$, $n/(n+1) = 80/81 = 0.987654321$, with the digits in descending order. The result is again "pan-digital" when this number is multiplied by $k = 1, 2, 4, 5, 7$ or 8. Interesting patterns result.

6. Among the values of $(10n/9)^2$, we have:

   (a) At $n = 3$, $(30/9)^2 = 11.1111111$; at $n = 6$, $(60/9)^2 = 44.44444444$; at $n = 30$, $(300/9)^2 = 1111.11111$.

   (b) At $n = 2$, $(20/9)^2 = 4.938271605$, where the even positions in ascending order are 0, 1, 2, 3, 4, and the odd positions in ascending order are 5, 6, 7, 8, 9.
   At $n = 4$, $(40/9)^2 = 19.75308642$, where the odd digits precede the even digits.
   At $n = 8$, $(80/9)^2 = 79.01234568$, a cycling of 0 through 9 with 8 out of place.
   At $n = 13$, $(130/9)^2 = 208.6419753$, where the even digits precede the odd digits.
   At $n = 14$, $(140/9)^2 = 241.9753086$, where the odd digits (19753) are flanked by the even digits.
   At $n = 20$, $(200/9)^2 = 49.38271605$, the same pattern as at $n = 2$.
   At $n = 26$, $(260/9)^2 = 834.5679012$, a cycling of 0 through 9 with 8 out of place.

7. If rounded to the nearest digit, a/b with $0 < a < b < 30$ is never "pandigital". However, on my Radio Shack 10-Digit Scientific Calculator EC-4032, I read $5/19 = 0.263157894$.

8. $(2143/22)^{\frac{1}{4}} = 3.141592653$, the first ten digits of π.

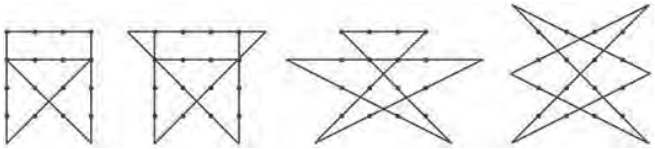
Reprinted from Vol. 57, No. 3, September 2007 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

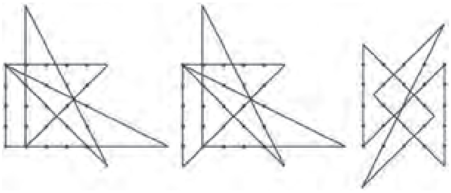
CONNECT THE DOTS

Solomon W. Golomb

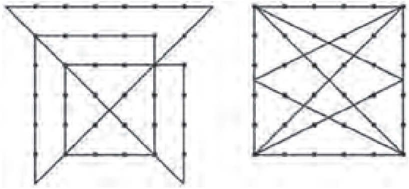
1. The four 6-segment circuits on the 4 × 4 array of dots are:



2. Here are three inequivalent 8-segment circuits on the 5 × 5 array of dots.

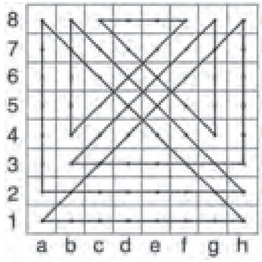


3. Here are two 10-segment circuits on the 6 × 6 array of dots.

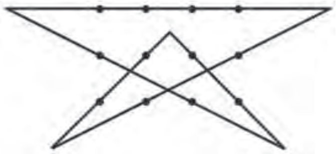


The solution on the right stays within the convex hull of the 6 × 6 array of dots.

4. This 14-move queen's tour of the chessboard was first published by Sam Loyd. It is included in *Sam Loyd and His Chess Problems*, compiled by Alain C. White, published 1913 by Whitehead and Miller; Dover reprint, 1962. The queen's circuit is: a1-h1-a8-a2-h2-b8-b4-f8-c8-g4-g8-b3-h3-h8-a1.



5. Here is the unique 5-segment circuit on the 3 × 4 array of dots.



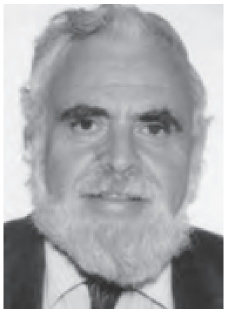
Reference. The definitive article on this subject is by S.W. Golomb and J.L. Selfridge, "Unicursal Polygonal Paths and Other Graphs on Point Lattices," *Pi Mu Epsilon Journal*, Fall, 1970.

Reprinted from Vol. 57, No. 4, December 2007 issue of Information Theory Newsletter

GOLOMB'S PUZZLE COLUMN™

EASY PROBABILITIES SOLUTIONS

Solomon W. Golomb



1. There are $\binom{10}{5} = 252$ ways to select 5 of the 10 decimal digits. When these are arranged in ascending order as a $< b < c < d < e$, the only way $a + b + c > d + e$ can occur is when the five selected numbers are {5, 6, 7, 8, 9}, so that 5+6+7 > 8+9. Thus, the probability of this occurring "at random" is only $\frac{1}{252} = 0.00396825\ldots$.

2. To maximize the probability that a green marble will be selected, place a single green marble in one jar, with the remaining n - 1 green marbles and all n red marbles in the second jar. If the contestant chooses the first jar (with probability $\frac{1}{2}$), the selected marble will be green. If the second jar is chosen, the probability of a green marble being drawn is $\frac{(n-1)}{(2n-1)}$. Thus the probability of a green marble being selected, for this arrangement, is $\frac{1}{2}(1 + \frac{n-1}{2n-1})$, which is $\frac{2}{3}$ if n has the minimum value $n = 2$, but tends to the limiting value of $\frac{3}{4}$ as n increases.

3. If North-South have all the hearts then East-West have none. Thus the probability that you and your partner (together) have all the hearts is the same as the probability that the two of you have none.

4. If A is a stronger player than B, your probability of winning two consecutive matches is better in the sequence ABA than in the sequence BAB. You can calculate this exactly if P_1 is the probability that A will beat you and P_2 is the probability that B will beat you, with $P_1 > P_2$. Intuitively, the result follows from: to win two matches (of the three) in a row, you *must* win the middle match, and B is weaker than A. Also, the sequence BAB gives you only *one* chance to defeat strong player A, while ABA gives you two chances.

5. (a) If *at least one* of the two children is a girl, there are three equally likely cases (in "birth order"): BG, GB, GG, and the probability that *both* are girls is $\frac{1}{3}$.

   (b) If *the older child* is a girl, the younger child is equally likely to be a boy or a girl; so in this case the probability that *both* are girls is $\frac{1}{2}$.

6. The marble originally in the jar is either black (B_1) or white (W_1). A new white marble (W_2) is inserted. The jar now contains either B_1W_2 or W_1W_2 (equally likely). When a marble is removed and observed to be white, there are now *three* equally likely cases: i) W_2 out, B_1 still in, ii) W_2 out, W_1 still in ; or iii) W_1 out, W_2 still in. (Observing the withdrawn marble to be white eliminated the case iv) B_1 out, W_2 still in.) Thus the probability that the unseen marble is white is $\frac{2}{3}$.

Reference. Problem 1 is my own, and previously unpublished. Some version of each of the remaining problems (all oldies) can be found in Martin Gardner's *Colossal Book of Short Puzzles and Problems,* W. Norton & Co. 2006, which identifies Problem 6 as coming from Lewis Carroll's book *Pillow Problems.*

In Memoriam: Mary Elizabeth (Betty) Moore Shannon

Reprinted from the Boston Globe (with permission of the Shannon family).

Mary Elizabeth (Betty) Moore Shannon, 95, formerly of Winchester, Massachusetts, died May 1, 2017 at her home at Brookhaven in Lexington, Massachusetts. She was born in New York City to Vilma Ujlaky Moore and James E. Moore.

Betty excelled academically in high school, winning a full scholarship to New Jersey College for Women (now Rutgers' Douglass College) where she graduated Phi Beta Kappa with a degree in mathematics. Upon graduation, she began working the next day at Bell Laboratories in Manhattan as a "computer," one of a group of women whose job was to do the mathematical calculations required by the engineers. She likened it to a secretarial pool for math majors. She was promoted to Technical Assistant at Bell, and worked with John Pierce, inventor of the communications satellite, collaborating on several projects including a Bell Labs Technical Memorandum entitled "Composing Music by a Stochastic Process."

At Bell Labs, Betty met Dr. Claude Shannon, the creator of Information Theory. They married in 1949 and were devoted to each other until Claude's death in 2001. She and Claude shared a playful sense of humor, and Betty assisted Claude in building some of his most famous inventions. She did much of the wiring of Theseus, the Maze-Solving Mouse, a pioneering experiment in artificial intelligence, and during a memorable trip to Las Vegas, helped test a device designed to beat the house at roulette, considered by many to be the first wearable computer.

Betty left Bell in 1951 to raise a family. She became an avid weaver, an interest she pursued for 40 years. She joined the Boston Weaver's Guild, served as Dean of the Guild from 1976-1978, and received the Guild's Distinguished Achievement Award. She worked closely with the Handweavers Guild of America for many years and received an honorary Life Membership in 1996. She was a member of the Cross Country Weavers and the Wednesday Weavers. In the '70s, Betty was one of the first explorers of computerized hand weaving, though she found, in the end, that she preferred the less technological approach.

In her later years she developed an interest in genealogy and never lost her love for all things mathematical.

Betty is survived by her son Andrew, her daughter Peggy, Peggy's partner Nina, and their daughters Nadja and Eva. Services will be private. Donations in her memory can be made to Associate Alumnae of Douglass College.

Nominations Sought for 2018 and 2019 Leaders

Volunteers needed to serve as corporate officers and committee chairs and members

By Howard E. Michel, Chair

2017 IEEE Nominations and Appointments Committee

IEEE is governed by volunteer members and depends on them for many things, including editing IEEE publications, organizing conferences, coordinating regional and local activities, authoring and authorizing publication of standards, leading educational activities, and identifying individuals for IEEE recognitions and awards.

The Nominations and Appointments (N&A) Committee is responsible for developing recommendations to be sent to the Board of Directors and the IEEE Assembly on staffing many volunteer positions including candidates for president-elect and corporate officers. Accordingly, the N&A Committee is seeking nominees for the leadership positions.

How to Nominate

For information about the positions, including qualifications and estimates of the time required by each position during the term of office, check the Guidelines for Nominating Candidates at www.ieee.org. To nominate a person for a position, complete the online IEEE Nominations & Appointments Committee Nomination Form.

Recent Publications

IEEE Transactions on Information Theory

Table of content for volumes 63(3), 63(4), 63(5).

Vol. 63(3): Mar. 2017.

SHANNON THEORY

T. J. Lim and M. Franceschetti	Information Without Rolling Dice	1349
D. Shaviv, A. Özgür, and H. H. Permuter	Capacity of Remotely Powered Communication	1364
O. Sabag, H. H. Permuter, and H. D. Pfister	A Single-Letter Upper Bound on the Feedback Capacity of Unifilar Finite-State Channels	1392
U. Pereg and I. Tal	Channel Upgradation for Non-Binary Input Alphabets and MACs	1410

CODING THEORY AND TECHNIQUES

F. L. Piñero	The Structure of Dual Schubert Union Codes	1425
L. Jin and C. Xing	New MDS Self-Dual Codes From Generalized Reed–Solomon Codes	1434
F. Jardel and J. J. Boutros	Edge Coloring and Stopping Sets Analysis in Product Codes With MDS Components	1439
G. Garrammone, D. Declercq, and M. P. C. Fossorier	Weight Distributions of Non-Binary Multi-Edge Type LDPC Code Ensembles: Analysis and Efficient Evaluation	1463
C. Rush, A. Greig, and R. Venkataramanan	Capacity-Achieving Sparse Superposition Codes via Approximate Message Passing Decoding	1476
I. Tal	On the Construction of Polar Codes for Channels With Moderate Input Alphabet Sizes	1501
K. Huang, U. Parampalli, and M. Xian	On Secrecy Capacity of Minimum Storage Regenerating Codes	1510

SPARSE RECOVERY, SIGNAL PROCESSING, LEARNING, ESTIMATION

T. Wadayama	Nonadaptive Group Testing Based on Sparse Pooling Graphs	1525
R. M. Castro and E. Tánczos	Adaptive Compressed Sensing for Support Recovery of Structured Sparse Sets	1535
C. Cartis and A. Thompson	Quantitative Recovery Conditions for Tree-Based Compressed Sensing	1555
A. Montanari, D. Reichman, and O. Zeitouni	On the Limitation of Spectral Methods: From the Gaussian Hidden Clique Problem to Rank One Perturbations of Gaussian Tensors	1572
A. Xu and M. Raginsky	Information-Theoretic Lower Bounds on Bayes Risk in Decentralized Estimation	1580
M. Genzel	High-Dimensional Estimation of Structured Signals From Non-Linear Observations With General Convex Loss Functions	1601
M. Kohler and A. Krzyżak	Nonparametric Regression Based on Hierarchical Interaction Models	1620

COMMUNICATION NETWORKS

N. Xie and S. Weber	Delay on Broadcast Erasure Channels Under Random Linear Combinations	1631
S.-W. Jeon, S.-N. Hong, M. Ji, G. Caire, and A. F. Molisch	Wireless Multihop Device-to-Device Caching Networks	1662
F. Yavuz, J. Zhao, O. Yağan, and V. Gligor	k-Connectivity in Random K-Out Graphs Intersecting Erdős-Rényi Graphs	1677

GAUSSIAN NETWORKS

J. M. Romero-Jerez and F. J. Lopez-Martinez	A New Framework for the Performance Analysis of Wireless Communications Under Hoyt (Nakagami-q) Fading	1693
F. Parvaresh and H. Soltanizadeh	Diversity-Multiplexing Trade-Off of Half-Duplex Single Relay Networks	1703
D. Castanheira, A. Silva, and A. Gameiro	Retrospective Interference Alignment: Degrees of Freedom Scaling With Distributed Transmitters	1721
Z. Zheng, L. Wei, R. Speicher, R. R. Müller, J. Hämäläinen, and J. Corander	Asymptotic Analysis of Rayleigh Product Channels: A Free Probability Approach	1731
L. V. Truong, S. L. Fong, and V. Y. F. Tan	On Gaussian Channels With Feedback Under Expected Power Constraints and With Non-Vanishing Error Probabilities	1746

SOURCE CODING

L. Zhou, V. Y. F. Tan, and M. Motani	Discrete Lossy Gray–Wyner Revisited: Second-Order Asymptotics, Large and Moderate Deviations	1766

QUANTUM INFORMATION THEORY

M. M. Wilde, M. Tomamichel, and M. Berta	Converse Bounds for Private Communication Over Quantum Channels	1792
M. Fukuda and G. Gour	Additive Bounds of Minimum Output Entropies for Unital Channels and an Exact Qubit Formula	1818
W. Kumagai and M. Hayashi	Second-Order Asymptotics of Conversions of Distributions and Entangled States Based on Rayleigh-Normal Probability Distributions	1829

SECURE COMMUNICATION

R. A. Chou, B. N. Vellambi, M. R. Bloch, and J. Kliewer	Coding Schemes for Achieving Strong Secrecy at Negligible Cost	1858
N. Shlezinger, D. Zahavi, Y. Murin, and R. Dabora	The Secrecy Capacity of Gaussian MIMO Channels With Finite Memory	1874
P. Mukherjee, J. Xie, and S. Ulukus	Secure Degrees of Freedom of One-Hop Wireless Networks With No Eavesdropper CSIT	1898
M. Mirmohseni and P. Papadimitratos	Secrecy Capacity Scaling in Large Cooperative Wireless Networks	1923

Vol. 63(4): Apr. 2017.

CODING THEORY AND TECHNIQUES

T. Johnsen and H. Verdure	Generalized Hamming Weights for Almost Affine Codes	1941
R. Saptharishi, A. Shpilka, and B. L. Volk	Efficiently Decoding Reed–Muller Codes From Random Errors	1954
V. Guruswami and C. Wang	Deletion Codes in the High-Noise and High-Rate Regimes	1961
C. Schoeny, A. Wachter-Zeh, R. Gabrys, and E. Yaakobi	Codes Correcting a Burst of Deletions or Insertions	1971
M. El-Khamy, H. Mahdavifar, G. Feygin, J. Lee, and I. Kang	Relaxed Polar Codes	1986
M. Ye and A. Barg	Explicit Constructions of High-Rate MDS Array Codes With Optimal Repair Bandwidth	2001
N. Raviv, N. Silberstein, and T. Etzion	Constructions of High-Rate Minimum Storage Regenerating Codes Over Small Fields	2015
H. Khodaiemehr and D. Kiani	Construction and Encoding of QC-LDPC Codes Using Group Rings	2039
Z. Wang, H. M. Kiah, Y. Cassuto, and J. Bruck	Switch Codes: Codes for Fully Parallel Reconstruction	2061
D. T. H. Kao, M. A. Maddah-Ali, and A. S. Avestimehr	Blind Index Coding	2076
Y.-C. Huang	Lattice Index Codes From Algebraic Number Fields	2098
S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi	Efficient Algorithms for Noisy Group Testing	2113

SPARSE RECOVERY, SIGNAL PROCESSING, LEARNING, ESTIMATION

M. Gavish and D. L. Donoho	Optimal Shrinkage of Singular Values	2137
S. Said, L. Bombrun, Y. Berthoumieu, and J. H. Manton	Riemannian Gaussian Distributions on the Space of Symmetric Positive Definite Matrices	2153
G. Puy, M. E. Davies, and R. Gribonval	Recipes for Stable Linear Embeddings From Hilbert Spaces to $\mathbb{R}^m$	2171
A. Kumar	On Bandlimited Field Estimation from Samples Recorded by a Location-Unaware Mobile Sensor	2188
C. Suh, V. Y. F. Tan, and R. Zhao	Adversarial Top-K Ranking	2201
E. Chaumette, A. Renaux, and M. N. El Korso	A Class of Weiss–Weinstein Bounds and Its Relationship With the Bobrovsky–Mayer-Wolf–Zakaï Bounds	2226

SHANNON THEORY

S. Hu and O. Shayevitz	The ρ-Capacity of a Graph	2241
X. Wu, A. Özgür, and L.-L. Xie	Improving on the Cut-Set Bound via Geometric Analysis of Typical Sets	2254
P. Moulin	The Log-Volume of Optimal Codes for Memoryless Channels, Asymptotically Within a Few Nats	2278
A. Xu and M. Raginsky	Information-Theoretic Lower Bounds for Distributed Function Computation	2314
A. Sharov and R. M. Roth	On the Capacity of Generalized Ising Channels	2338
B. T. Ávila and R. M. Campello de Souza	Meta-Fibonacci Codes:Efficient Universal Coding of Natural Numbers	2357

GAUSSIAN CHANNELS

H. Ghozlan and G. Kramer	Models and Information Rates for Wiener Phase Noise Channels	2376
S. Yang and S. Shamai	On the Multiplexing Gain of Discrete-Time MIMO Phase Noise Channels	2394
A. Ben-Yishai and O. Shayevitz	Interactive Schemes for the AWGN Channel with Noisy Feedback	2409

SEQUENCES

QUANTUM INFORMATION THEORY

COMPLEXITY AND CRYPTOGRAPHY

SECURE COMMUNICATION

COMMENTS AND CORRECTIONS

Vol. 63(5): May. 2017.

SHANNON THEORY

CODING THEORY AND TECHNIQUES

SOURCE CODING

SPARSE RECOVERY, SIGNAL PROCESSING, LEARNING, ESTIMATION

COMMUNICATION NETWORKS

GAUSSIAN NETWORKS

QUANTUM INFORMATION THEORY

COMMENTS AND CORRECTIONS

Problems of Information Transmission Volume 53, Issue 1, January 2017

Some “goodness” properties of LDA lattices
S. Vatedka, N. Kashyap
Pages 1–29

Bounds on the rate of separating codes
I. V. Vorob’ev
Pages 30–41

Optimal conflict-avoiding codes for 3, 4 and 5 active users
T. Baicheva, S. Topalova
Pages 42–50

Remark on balanced incomplete block designs, near-resolvable block designs, and q-ary constant-weight codes
L. A. Bassalygo, V. A. Zinoviev
Pages 51–54

Linear algorithm for minimal rearrangement of structures
K. Yu. Gorbunov, V. A. Lyubetsky
Pages 55–72

Model of a random geometric graph with attachment to the coverage area
S. N. Khoroshenkikh, A. B. Dainiak
Pages 73–83

Occurrence numbers for vectors in cycles of output sequences of binary combining generators
O. V. Kamlovskii
Pages 84–91

Number of curves in the generalized Edwards form with minimal even cofactor of the curve order
A. V. Bessalov, O. V. Tsygankova
Pages 92–101

5th International Castle Meeting on Coding Theory and Applications

PRELIMINARY CALL FOR PAPERS

This is the first announcement of the Fifth International Castle Meeting on Coding Theory and Applications (5ICMCTA), which will take place in Vihula Manor, Estonia, from Monday, August 28th, to Thursday, August 31st, 2017. Information about the 5ICMCTA can be found at http://www.castle-meeting-2017.ut.ee/ .

We solicit submissions of previously unpublished contributions related to coding theory, including but not limited to the following areas: *Codes and combinatorial structures, Algebraic-geometric codes, Network coding, Codes for storage, Quantum codes, Convolutional codes, Codes on graphs, Iterative decoding, Coding applications to cryptography and security, Other applications of coding theory*.

Organization:

General chair: *Vitaly Skachek*

Scientific Committee co-chairs: *Ángela Barbero* and *Øyvind Ytrehus*

Publicity: *Yauhen Yakimenka*

Scientific Committee

Alexander Barg • Irina Bocharova • Eimear Byrne • Joan-Josep Climent • Gerard Cohen • Olav Geil • Marcus Greferath • Tor Helleseth • Tom Høholdt • Camilla Hollanti • Kees S. Immink • Frank Kschischang • Boris Kudryashov • San Ling • Daniel Lucani • Gary McGuire • Sihem Mesnager • Muriel Médard • Diego Napp • Frederique Oggier • Patric Östergard • Raquel Pinto • Paula Rocha • Joachim Rosenthal • Eirik Rosnes • Moshe Schwartz • Vladimir Sidorenko • Patrick Sole • Leo Storme • Rüdiger Urbanke • Pascal Vontobel • Dejan Vukobratovic • Jos Weber • Gilles Zémor

Important dates:

Paper submission:	May 1, 2017
Notification of decision:	June 12, 2017
Final version paper submission:	July 3, 2017

The 10th International Workshop on Coding and Cryptography

WCC 2017

Saint-Petersburg, Russia, September 18–22, 2017

http://wcc2017.suai.ru/

ANNOUNCEMENT AND CALL FOR PAPERS

Organizing committee:

- Pierre Loidreau (**co-chair**, DGA, U. Rennes 1, France)
- Evgeny Krouk (**co-chair**, SUAI, Russia)
- Veronika Prokhorova (SUAI, Russia)
- Evgeny Bakin (SUAI, Russia)

Local organization:

- Oksana Novikova
- Maria Shelest

Invited Speakers:

- Alexander Barg (U. Maryland, USA)
- Claude Carlet (U. Paris 8 and Paris 13, France)
- Camilla Hollanti (Aalto U., Finland)
- Grigory Kabatiansky (Skoltech and IITP RAS, Moscow, Russia)
- Patric Östergård (Aalto U., Finland)

This is the tenth in the series of biannual workshops Coding and Cryptography. It is organized by INRIA, SUAI and Skoltech and will be held in the main building of SUAI (http://suai.ru/), Saint-Petersburg, Russia.

Conference Themes. Our aim is to bring together researchers in all aspects of coding theory, cryptography and related areas, theoretical or applied.

Topics include, but are not limited to:

- coding theory: error-correcting codes, decoding algorithms, related combinatorial problems;
- algorithmic aspects of cryptology: symmetric cryptology, public-key cryptography, cryptanalysis;
- discrete mathematics and algorithmic tools related to these two areas, such as: Boolean functions, sequences, finite fields, related algebraic systems.

Submissions. Those wishing to contribute a talk are invited to submit a 6-10 page extended abstract, before April 6, 2017 (23:59 Greenwich). The submission server is now open, information on the submission process is available at http://wcc2017.suai.ru/submission.html.

Full papers. After the conference, authors of accepted abstracts will be invited to submit a full paper for the proceedings to appear as a special issue of the journal "Designs Codes and Cryptography". Contributions will be thoroughly refereed.

Important dates (for extended abstracts)**:**

- **Submission by April 6, 2017**
- **Notification by May 24, 2017**
- **Final version by June 26, 2017**

Program committee:

- Daniel Augot (**co-chair**, INRIA, France)
- Delphine Boucher (U. Rennes 1, France)
- Lilya Budaghyan (U. Bergen, Norway)
- Eimear Byrne (UC Dublin, Ireland)
- Pascale Charpin (INRIA, France)
- Alain Couvreur (INRIA, France)
- Ilia Dumer (UC Riverside, USA)
- Tuvi Etzion (CSD Technion, Israel)
- Markus Grassl (MPL Erlangen, Germany)
- Tor Helleseth (U. Bergen, Norway)
- Thomas Honold (U. Zhejiang, China)
- Thomas Johansson (Lund U., Sweden)
- Gohar Khyureghan (U. Magdeburg, Germany)
- Ivan Landjev (NBU Sofia, Bulgaria)
- Subhamoy Maitra (ISI Kolkata, India)
- Ryutaroh Matsumoto (Tokyo Tech., Japan)
- Gary McGuire (UC Dublin, Ireland)
- Sihem Mesnager (U. Paris 8 and Paris 13, France)
- Marine Minier (INSA-Lyon, France)
- Kaisa Nyberg (AU Helsinki, Finland)
- Ayoub Otmani (U. Rouen-Normandie, France)
- Ferruh Ozbudak (METU Ankara, Turkey)
- Kevin Phelps (AU Auburn, USA)
- Alexander Pott (U. Magdeburg, Germany)
- Josep Rifa (UA Barcelona, Spain)
- Palash Sarkar (ISI Kolkata, India)
- Natalia Shekhunova (SUAI St.Petersburg, Russia)
- Vladimir Sidorenko (TU Munich, Germany)
- Faina I. Solov'eva (**co-chair**, IM Sobolev, Russia)
- Jean-Pierre Tillich (INRIA, France)
- Alev Topuzoğlu (Sabanci U. Istanbul, Turkey)
- Peter Trifonov (PU St.Petersburg, Russia)
- Michail Tsfasman (CNRS and IITP RAS, Marseille, France)
- Serge Vladuts (Aix-Marseille U. and IITP RAS, France)
- Arne Winterhof (Austrian Acad. of Sc., Linz)
- Gilles Zémor (U. Bordeaux, France)
- Victor Zinoviev (IITP RAS, Moscow, Russia)
- Victor Zyablov (IITP RAS, Moscow, Russia)

55th Allerton Conference

Call for Papers: Due July 10, 2017

Manuscripts can be submitted during June 16-July 10, 2017 with the submission deadline of July 10th being firm. Please follow the instructions at allerton.csl.illinois.edu.

CONFERENCE CO-CHAIRS: Naira Hovakimyan & Negar Kiyavash

INFORMATION FOR AUTHORS: Regular papers suitable for presentation in twenty minutes are solicited. Regular papers will be published in full (subject to a maximum length of eight 8.5" x 11" pages, in two column format) in the Conference Proceedings. Only papers that are actually presented at the conference and uploaded as final manuscripts can be included in the proceedings, which will be available after the conference on IEEE Xplore. For reviewing purposes of papers, a title and a five to ten page extended abstract, including references and sufficient detail to permit careful reviewing, are required.

IMPORTANT DATES - 2017

JULY 10 — Submission Deadline

AUGUST 7 — Acceptance Date *Authors will be notified of acceptance via e-mail by August 7, 2017, at which time they will also be sent detailed instructions for the preparation of their papers for the Conference Proceedings.*

AFTER AUGUST 7 — Registration Opens

OCTOBER 3-6 — Conference Dates

OCTOBER 3 — Opening Tutorial Lectures
Coordinated Science Lab, University of Illinois at Urbana-Champaign

OCTOBER 4-6 — Conference Sessions (Plenary Lecture October 6)
University of Illinois Allerton Park & Retreat Center: The Allerton House is located twenty-six miles southwest of the Urbana-Champaign campus of the University of Illinois in a wooded area on the Sangamon River. It is part of the fifteen-hundred acre Robert Allerton Park, a complex of natural and man-made beauty designated as a National natural landmark. Allerton Park has twenty miles of well-maintained trails and a living gallery of formal gardens, studded with sculptures collected from around the world.

OCTOBER 8 — Final Paper Deadline *Final versions of papers that are presented at the conference must be submitted electronically in order to appear in the Conference Proceedings and IEEE Xplore.*

All speaker information will be added when confirmed.

PAPERS PRESENTING ORIGINAL RESEARCH ARE SOLICITED IN THE AREAS OF:

- Biological information systems
- Coding techniques & applications
- Coding theory
- Data storage
- Information theory
- Multiuser detection & estimation
- Network information theory
- Sensor networks in communications
- Wireless communication systems
- Intrusion/anomaly detection & diagnosis
- Network coding
- Network games & algorithms
- Performance analysis
- Pricing & congestion control
- Reliability, security & trust
- Decentralized control systems
- Robust & nonlinear control
- Adaptive control & automation
- Robotics
- Distributed & large-scale systems
- Complex networked systems
- Optimization
- Dynamic games
- Machine learning & learning theory
- Signal models & representations
- Signal acquisition, coding, & retrieval
- Detection & estimation
- Learning & inference
- Statistical signal processing
- Sensor networks
- Data analytics

WEBSITE | allerton.csl.illinois.edu
EMAIL | amellis@illinois.edu

ECE ILLINOIS
Department of Electrical and Computer Engineering

CSL: COORDINATED SCIENCE LAB
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

ALLERTON CONFERENCE

2017 IEEE Information Theory Workshop

Kaohsiung Exhibition Center, Kaohsiung, Taiwan / **November 6–10, 2017**

http://www.itw2017.org

General Co-Chairs
Po-Ning Chen (NCTU, Taiwan)
Gerhard Kramer (TUM, Germany)
Chih-Peng Li (NSYSU, Taiwan)

Technical Program Co-Chairs
Hsiao-feng (Francis) Lu (NCTU, Taiwan)
Stefan M. Moser (ETHZ, Switzerland)
Chih-Chun Wang (Purdue Univ., USA)

Finances
Chung-Hsuan Wang (NCTU, Taiwan)
Jwo-Yuh Wu (NCTU, Taiwan)

Publicity
Osvaldo Simeone (NJIT, USA)
Hsaun-Jung Su (NTU, Taiwan)

Publications
Stefano Rini (NCTU, Taiwan)
I-Hsiang Wang (NTU, Taiwan)

Local Arrangement
Celeste Lee (NSYSU, Taiwan)
Fan-Shuo Tseng (NSYSU, Taiwan)
Chao-Kai Wen (NSYSU, Taiwan)

Webmaster
Yao-Win (Peter) Hong (NTHU, Taiwan)

International Advisory Committee
Toru Fujiwara (Osaka Univ., Japan)
Shuo-Yen (Robert) Li (CUHK, China)
Alon Orlitsky (UCSD, USA)
A. J. Han Vinck (UDE, Germany)
Muriel Médard (MIT, USA)
Andrea Goldsmith (Stanford, USA)
Raymond W. Yeung (CUHK, China)

Sponsors
IEEE Information Theory Society

Co-Sponsors
National Chiao-Tung University
National Sun Yat-sen University

The 2017 IEEE Information Theory Workshop will take place in Kaohsiung, Taiwan, from November 6 to 10, 2017. Based at the southern tip of the island and facing the Taiwan Strait, Kaohsiung is the second largest city in Taiwan and is one of the major seaports in Asia. Love-River cruises, night markets, delicious seafood, and day tours out into nature or to historic Tainan are some of the many attractions awaiting you in Southern Taiwan. Situated directly by the waterfront, the Kaohsiung Exhibition Center (KEC) serves as workshop venue. It is a brand-new and multi-functional facility, designed by an international, pro-environment team of architects and built in the shape reminding of a billowing sail. The workshop participants will have an unforgettable experience visiting and enjoying some of the most dazzling attractions in Kaohsiung.

Call for Papers

Interested authors are encouraged to submit previously unpublished contributions in all areas of information theory with special emphasis on the following :

- **Information Theory for Content Distribution**
  - Distributed data storage
  - Peer-to-peer network coded broadcasting
  - Coded caching for wireless and wireline transmissions
  - Delay-constrained communications
- **Information Theory and Biology**
  - Information theory and intercellular communications
  - Information theory and neuroscience
  - Information-theoretical analysis of biologically-inspired communication systems
- **Information Theory and Quantum Communications**
  - Quantum information
  - Quantum computation
  - Quantum cryptography
- **Information Theory and Coding for Memories**
  - New coding techniques for non-volatile memory channels
  - Coding and signal processing for dense memory
  - Multi-dimensional coding for storage channels

Paper Submission

Paper submission guidelines will be posted on the workshop's website : **http://www.itw2017.org**

Poster

The technical program will feature a poster session. Details about poster submissions will be announced on the workshop's website by late July, 2017.

Important Dates

Paper submission deadline : **May 7, 2017**
Acceptance notification : **July 21, 2017**

Call For Papers

Call for Papers

2018 International Zurich Seminar on Information and Communication

February 21 – 23, 2018

The 2018 International Zurich Seminar on Information and Communication will be held at the Hotel Zürichberg in Zurich, Switzerland, from Wednesday, February 21, through Friday, February 23, 2018. High-quality original contributions of both applied and theoretical nature are solicited in the areas of:

- Wireless Communication
- Information Theory
- Coding Theory and its Applications
- Detection and Estimation
- Data Storage
- Optical Communication
- Fundamental Hardware Issues
- Network Algorithms and Protocols
- Network Information Theory and Coding
- Cryptography and Data Security

Invited speakers will account for roughly half the talks. In order to afford the opportunity to learn from and communicate with leading experts in areas beyond one's own specialty, no parallel sessions are anticipated. All papers should be presented with a wide audience in mind.

Papers will be reviewed on the basis of a manuscript (A4, not exceeding 5 pages) of sufficient detail to permit reasonable evaluation. Authors of accepted papers will be asked to produce a manuscript not exceeding 5 pages in A4 double column format that will be published in the Proceedings. Authors will be allowed twenty minutes for presentation.

The deadline for submission is **September 17, 2017**.

Additional information will be posted at

`http://www.izs.ethz.ch/`

We look forward to seeing you at IZS.

Amos Lapidoth and Stefan M. Moser, Co-Chairs.

Conference Calendar

DATE	CONFERENCE	LOCATION	WEB PAGE	DUE DATE
June 6–9, 2017	**2017 North-American School of Information Theory**	Atlanta, Georgia, USA	http://www.itsoc.org/conferences/schools/2017-north-american-school-of-information-theory	Passed
June 11–14, 2017	**15th Canadian Workshop on Information Theory**	Quebec City, Quebec, Canada	http://cwit.ca/2017/	Passed
June 25–30, 2017	**IEEE International Symposium on Information Theory (ISIT)**	Aachen, Germany	http://www.isit2017.org	Passed
July 3, 2017	**Munich Workshop on Coding and Applications**	Munich, Germany	http://www.lnt.ei.tum.de/en/events/munich-workshop-on-coding-and-applications-2017-mwca2017	Passed
July 3–6, 2017	**The 18th IEEE International Workshop on Signal Processing Advances in Wireless Communications**	Sapporo, Japan	http://www.spawc2017.org/public.asp?page=home.html	Passed
July 10–14, 2017	**Croucher Summer Course in Information Theory**	Chinese University of Hong Kong	http://cscit.ie.cuhk.edu.hk/	Passed
August 28–31, 2017	**Fifth International Castle Meeting on Coding Theory and Applications (5ICMCTA)**	Vihula Manor, Estonia	http://www.castle-meeting-2017.ut.ee/	Passed
September 11–22, 2017	**Quasi-Cyclic and Related Algebraic Codes**	Ankara, Turkey	http://www.isit2017.org	June 11
September 18–22, 2017	**The Tenth International Workshop on Coding and Cryptography 2017**	Saint-Petersburg, Russia	http://wcc2017.suai.ru/	Passed
October 3–6, 2017	**55th Annual Allerton Conference on Communication, Control, and Computing**	University of Illinois at Urbana-Champaign	http://allerton.csl.illinois.edu/	July 10, 2017
October 15–17, 2017	**58th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2017)**	Berkeley, California, USA.	http://focs17.simons.ber/keley.edu	Passed
November 6–10, 2017	**IEEE Information Theory Workshop**	Kaohsiung, Taiwan	http://www.itw2017.org/	Passed
December 4–8, 2017	**IEEE GLOBECOM**	Singapore	http://globecom2017.ieee-globecom.org/	Passed
February 21–23, 2018	**2018 International Zurich Seminar on Information and Communication**	Zurich, Switzerland	http://www.izs.ethz.ch/	September 17, 2017

Major COMSOC conferences: http://www.comsoc.org/confs/index.html