WASHINGTON
STATISTICAL
SOCIETY

WSS NEWS

June 2008

Annual Dinner

This year's WSS Annual Dinner is at the Meiwah restaurant in Chevy Chase, Maryland (Friendship Heights Metro), Wednesday, June 25, 2008. The Gertrude Cox Award winner is Dr. Thomas Lumley from the University of Washington. Dr Lumley will speak at the dinner about open source statistical software. The price for the dinner is $45 person.

Thomas Lumley: "Open source statistical software: how, why, where?"

Over the current decade the R statistics environment and its package system have gone from being too obscure to be worth citing to being too well-known to need citing. I will talk about some current and historical issues in the design and development of R and of my R 'survey' package. I will also discuss how open-source statistical software fits (or fails to fit) the standard explanations for the success of the open-source model.

The registration flyer is attached.

WSS and Other Seminars

(All events are open to any interested persons)

June

3 Tues. **New Methods in Network And Spatial Sampling**

10 Tues. **Nonresponse Adjustments in Survey Applications**

11 Wed. **Save the Date:** *Elections and Exit Polls*

17 Tues. **Recent Developments in Address-based Sampling**

18 Wed. **Entropy and ROC Based Methods for SNP Selection and GWA Study**

30 Mon. **Statistical Policy Issues Arising in Carrying Out the Requirements of the Prison Rape Elimination Act of 2003**

Also available on the Web at the following URL: http://www.scs.gmu.edu/~wss/

Announcement

SIGSTAT Topics for Spring 2008

June 18, 2008: Survival Models in SAS: PROC PHREG – Part 3

Continuing the series of talks based on the book "Survival Analysis Using the SAS System: A Practical Guide" by Paul Allison begun in October 2007, we'll continue with Chapter 5: Estimating Cox Regression Models with PROC PHREG.

Topics covered are: Time-Dependent Covariates

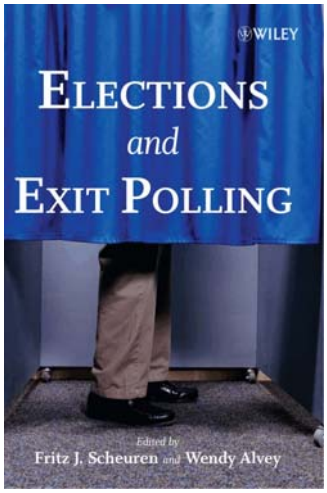
SIGSTAT is the Special Interest Group in Statistics for the **CPCUG**, the Capital PC User Group, and **WINFORMS**, the Washington Institute for Operations Research Service and Management Science.

All meetings are in Room S3031, 1800 M St, NW from **12:00 to 1:00**. Enter the South Tower & take the elevator to the 3rd floor to check in at the guard's desk. First-time attendees should contact Charlie Hallahan, 202-694-5051, hallahan@ers.usda.gov, and leave their name. Directions to the building & many links of statistical interest can be found at the **SIGSTAT** website, **http://www.cpcug.org/user/sigstat/**.

Announcement

Book Signing at Reiter's Books
1990 K Street NW, Washington DC
Wednesday, June 11th 12:00 noon-2:00pm

Lite refreshments (wine, cheese, and soft drinks) will be provided.



Wendy Alvey and Fritz Scheuren
Co-editors of Elections and Exit Polls

The newly-released book -- *Elections and Exit Polling*, edited by Fritz Scheuren and Wendy Alvey -- is a tribute to Warren Mitofsky, the father of exit polling. The volume, just published by John Wiley & Sons, Inc., consists of excerpts from interviews with Mitofsky shortly before he died and selected readings from recent statistical research related to election polling and exit polling.

Combining wisdom from one of the most notable names in the field along with findings from modern research and insightful recommendations for future practices, *Elections and Exit Polling* is an excellent supplement for political science and survey research courses at the upper-undergraduate or graduate level. It is also a one-of-a-kind reference for pollsters, survey researchers, statisticians, and anyone with a general interest in the methods behind global elections and exit polling.

All royalties from the book have been donated to the Warren J. Mitofsky Award for Excellence in Survey Research. This award is being co-sponsored by the American Association for Public Opinion Research and the American Statistical Association and is being managed by The Roper Center for Public Opinion Research. The award recognizes outstanding research or reporting of public opinion or survey methodology, especially work based on data from The Roper Center's public opinion archives.

The chapter titles are:

1. Introduction, by Fritz Scheuren and Wendy Alvey
2. The Infamous 2000 Election
3. 2004: Fraudulent Election?
4. Midterm Elections: 2006
5. Globe-Trotting Consultant
6. Looking Ahead: Recommendations for 2008 and Beyond
7. Technical Appendix

Sponsor: WSS Human Rights Section and WSS Public Policy Section

Program Announcement

Title: **New Methods in Network And Spatial Sampling**

Speaker: Steven K. Thompson, Simon Fraser University

Chair: Myron Katzoff, CDC/National Center for Health Statistics

Date/Time: Tuesday, June 3, 2008 / 12:30 – 2:00 p.m.

Location: Bureau of Labor Statistics Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics, you need to email your name, affiliation and seminar name to wss_seminar@bls.gov (note that there is an underscore after 'wss') by noon at least two days in advance of the seminar or call 202-691-7524 and leave a message with this information. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Ave., NE. Take the Red Line to Union Station.

Sponsor: Defense and National Security Section

Abstract: Over the last several decades new sampling methods have been developed in response to problems in studies of difficult-to-sample populations and to theory on optimal designs. In this talk I'll describe a handful of such methods for sampling of populations in network and spatial settings.

Program Announcement

Title: **Nonresponse Adjustments in Survey Applications**

Chair: Nancy Bates, U.S. Census Bureau

Speakers: Frauke Kreuter, Joint Program in Survey Methodology, University of Maryland
Trena Ezzati-Rice, Agency for Healthcare Research and Quality

Discussant: Keith Rust, Westat and JPSM

Date/Time: Tuesday, June 10, 2008 / 12:30 - 2:00 p.m.

Location: Bureau of Labor Statistics, Conference Center. To be placed on the seminar attendance list at the Bureau of Labor Statistics you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least two days in advance of the seminar, or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.

Sponsor: WSS Methodology Program

Abstract (Kreuter):

Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Nonresponse: Examples from Multiple Surveys Nonresponse weighting is a commonly used method to adjust for bias due to unit nonresponse in surveys. Theory and simulations show that, in order to effectively reduce bias without increasing variance, a covariate used for nonresponse weighting adjustment needs to be highly associated with both response and the survey outcome. In practice, these requirements pose a challenge that is often overlooked. Recently some surveys have begun collecting supplementary data, such as interviewer observations and other proxy measures of key survey outcomes. These variables are promising candidates for nonresponse adjustment because they should be highly correlated with the actual outcomes. In the present study, we examine the extent to which traditional covariates and new proxy measures satisfy the weighting requirements for the National Survey of Family Growth, the Medical Expenditure Survey, the U.S. National Election Survey, the European Social Surveys and the University of Michigan Transportation Research Institute Survey. We provide empirical estimates of the association between proxy measures and the likelihood of response as well as the actual survey responses. We also compare unweighted and weighted estimates under various nonresponse models. Results show the difficulty of finding suitable covariates and the need to improve the quality of proxy measures.

Abstract (Ezzati-Rice):

Assessment of the Impact of Health Variables on Nonresponse Adjustment in the Medical Expenditure Panel Survey The Medical Expenditure Panel Survey (MEPS) is a large complex sample survey, designed to provide nationally representative annual estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. A new panel of households is selected each year for the MEPS from households that responded to the previous year's National Health Interview Survey (NHIS). Nonresponse is a

common problem in household sample surveys. To compensate for nonresponse and to reduce the potential bias of the survey estimates, two separate nonresponse adjustments are performed in development of analytic weights in MEPS. The first, the focus of this presentation, is an adjustment for dwelling unit (DU) level nonresponse to account for nonresponse among those households subsampled from NHIS for the MEPS. The adjustment is carried out using socio-economic, demographic, and health variables that are available for both respondents and nonrespondents. In this study, we examine the impact of health variables on the MEPS DU level nonresponse weight adjustment. Response propensity scores are calculated based on logistic regression models and quintiles of the propensity scores are used to adjust the MEPS base weights. Comparisons of the nonresponse adjusted weights and selected survey variables with and without inclusion of health variables as a nonresponse adjustment covariate are discussed.

Program Announcement

Title: **Recent Developments in Address-based Sampling**

Chair: Meena Khare, NCHS

Speakers: Mansour Fahimi, Marketing Systems Group

Discussant: Sylvia Dohrmann, Westat

Date/Time: Tuesday, June 17, 2008 / 12:30 - 2:00 p.m.

Sponsor: WSS Methodology Program

Location: Bureau of Labor Statistics, Conference Center Room 8

Attendance: In order to be placed on the seminar list attendance list at the Bureau of Labor Statistics you need to e-mail your name, affiliation, and seminar name to wss_seminar@bls.gov (underscore after 'wss') by noon at least 2 days in advance of the seminar or call 202-691-7524 and leave a message. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Avenue, NE. Use the Red Line to Union Station.

Abstract: Increasingly, survey researchers are reverting back to address-based methodologies to reach the general public for survey administration and related commercial applications. Essentially, there are three main factors for this change: evolving coverage problems associated with telephone-based methods; eroding rates of response to telephone contacts; and on the other hand, recent improvements in the databases of household addresses available to researchers. This presentation provides an assessment of these three factors along with an overview of the structure of the Delivery Sequence File (DSF) of the USPS that is often used for construction of address-based sampling frames. Moreover, key enhancements available for the DSF will be discussed. While reducing undercoverage bias particularly in rural areas where more households rely on P.O. Boxes and inconsistent address formats such enhancements enable researcher to develop more efficient sample designs as well as broaden their analytical possibilities through an expanded set of covariates for hypothesis testing and statistical modeling tasks.

Program Announcement

Title: **Entropy and ROC Based Methods for SNP Selection and GWA Study**

Speaker: Prof. Zhenqiu Liu
Department of Epidemiology and Preventive Medicine University of Maryland School of Medicine

Date/Time: Wednesday, June 18, 2008 / 11:00 a.m. to 12:00 p.m.

Location: Conference Room 9201. Two Rockledge Center, 6701 Rockledge Drive, Bethesda, MD 20892

Sponsor: NCI/DPPS/OBR

Abstract: GWA studies have become an important approach in the last few years as a means to elucidate associations between particular alleles and a predisposition to disease. Genome wide SNP selection and association study with entropy related methods has been proven to be useful in the literature. In this talk, we introduce a multi-locus LD measure with generalized mutual information. SNP tagging, genetic mapping, and association study are performed with the proposed LD measure and Monte Carlo methods. We also briefly introduce a ROC based statistical learning approach for SNP selection and association study and discuss methods to detect the rare alleles associated with disease

Program Announcement

Title: **Statistical Policy Issues Arising in Carrying Out the Requirements of the Prison Rape Elimination Act of 2003**

Speaker: Allen Beck, Bureau of Justice Statistics, U.S. Department of Justice

Discussant: Hermann Habermann, former Director, U.N. Statistics Division and former Deputy Director, U.S. Census Bureau

Chair: Shelly Wilkie Martinez, Office of Statistical and Science Policy, U.S. Office of Management and Budget

Date/Time: Monday, June 30, 2008, 12:30 – 2:00 p.m.

Location: Bureau of Labor Statistics Conference Center. To be placed on the seminar attendance list at the Bureau of labor Statistics, you need to email your name, affiliation and seminar name to wss_seminar@bls.gov (note that there is an underscore after 'wss') by noon at least two days in advance of the seminar or call 202-691-7524 and leave a message with this information. Bring a photo ID to the seminar. BLS is located at 2 Massachusetts Ave., NE. Take the Red Line to Union Station.

Sponsor: WSS Section on Public Policy

Abstract: Our speaker will discuss how the Bureau of Justice Statistics has approached its responsibilities under the Prison Rape Elimination Act of 2003. The act provides fairly detailed sampling specifications and requires BJS to publish prison- and jail-level data on the incidence of rape, and to identify the three "best" and "worst" of each. Given the sensitive nature of the content, the developmental nature of the data collections, and the administrative and enforcement purposes to which the data will be put, BJS has had to step carefully to maintain its position as the Justice Department's principal *statistical* agency. Our discussant will assess the unfolding BJS experience as a case study of agency practice against professional practice and ethical criteria embodied in frameworks such as the United Nation's *Fundamental Principles of Official Statistics* and in *Principles and Practices of a Statistical Agency*, a seminal publication of the Committee on National Statistics.

Announcement

The Statistics Department at The George Washington University will offer the following Graduate Courses in Fall 2008 (September 2 – December 20, 2008) at the main campus.

Enhance your statistical analysis skills by taking one or more of these courses. Registering as a non-degree student is easy - please visit www.gwu.edu/~regweb for relevant information.

For questions or further information please contact Dr. Reza Modarres, e-mail: reza@gwu.edu, ph: 202-994-6888.

Statistics 201-10. Mathematical Statistics. Thursday, 6:10pm-8:40pm.
Instructor: Dr. H. Mahmoud.

This is the first part of a two-part graduate level series in Mathematical Statistics. The objective of the course is to introduce students to the concepts of probability that are useful for understanding statistical theory (the course continues on to Stat 202 in Spring, which deals with the theory of statistical inference). Topics to be covered in Stat 201 include basics of probability theory (including conditional probability, Bayes theorem, random variables, density and mass functions), univariate transformations, expected value, moment generating function, common probability distributions (including binomial, normal, uniform), multivariate distributions and transformations, covariance, inequalities and sampling distributions. This is roughly chapters 1 through 5 of the text: *Statistical Inference* (2nd Ed.) by Casella, G. and Berger, R. L.; Duxbury Press, CA.

This course is required for MS and Ph.D. students in Statistics, and Biostatistics, and Ph.D. students in Epidemiology. Students from other quantitative fields such as Economics, Finance, Engineering etc. would also find the course very useful and are encouraged to join. Prerequisites: Multivariable Calculus (Math 33), and Linear Algebra (Math 124) or equivalent.

Statistics 207. Methods of Statistical Computing I. Tuesday, 06:10pm-08:40pm.
Instructor: Dr. Y. Lai.

Computing is essential for the practice of statistics. This course will introduce basic computational methods from a statistical point of view. In particular, the following general areas will be covered: (i) Fundamental of computer science; (ii) Numerical analysis and computer intensive methods; and (iii) Statistical computing and graphics.

Prerequisites include knowledge of a programming language, a course in matrix algebra and mathematical statistics.

Textbooks: Statistical Computing, by W. J. Kennedy and J. E. Gentle and
An Introduction to the Bootstrap, by B. Efron and R. J. Tibshirani.

Statistics 215. Applied Multivariate Analysis. Monday, 06:10pm-08:40pm.
Instructor: Dr. R. Modarres.

This course is intended for students interested in statistical analysis of several variables, most likely dependent, following a joint normal distribution. It covers inferential and descriptive multivariate techniques, including the multivariate normal distribution, assessing the assumption of normality, transformations to near normality, Hotelling test for the mean vector, confidence regions and simultaneous comparisons of component means, missing observations and the EM algorithm and principal components analysis. In addition to the text, other topics from the literature, including some non-parametric techniques will be covered. For each technique, the theoretical foundation is developed and applied to observations from behavioral, social, medical, and physical sciences. The computational aspects will include use of matrix algebra tools (SAS/IML). Prerequisites include a course in matrix algebra and mathematical statistics.

Textbook: Applied Multivariate Analysis, 6th Ed., by R.A. Johnson and D.W. Wichern.

Stat 217: Design of Experiments. Tuesday, 6:10pm-8:40pm.
Instructor: Dr. Z. Li.

This course is a graduate level introduction to Design of Experiments, an area of statistics concerned with the planning of scientific investigation. The main components of an experimental design are the selection of the independent and dependent variables to be studied, determination of sample size, and allocation of experimental units to experimental treatments.

Specific topics which will be covered in detail include Replication, Blocking, Randomization, Factorial and Fractional -Factorial experiments, Repeated Measures designs, and Latin Square designs. Prerequisite: Stat 157-58; Math 124.

Statistics 227. Survival Analysis. Wednesday, 6:10pm-8:40pm.
Instructor: Dr. Q. Pan.

This course will discuss parametric and nonparametric methods for the analyses of events observed in time (survival data). Topics include: survival distributions, Kaplan-Meier estimate of survival functions, Greenwood's formula, Mantel-Haenszel test, logrank and generalized logrank tests, Cox proportional hazards model, parametric regression models, and power and sample size calculations for survival analysis.

Prerequisite: Stat 201-2 or permission of instructor.

Statistics 257. Probability. Wednesday, 6:10pm-8:40pm.
Instructor: Dr. H. Mahmoud

This course will discuss rigorous modern measure-theoretic probability. No prior knowledge of measure theory is assumed; the necessary concepts will be developed as necessary. Topics to be covered include: Sigma fields and Probability measures, Probability Axioms, Lebesgue integration and expectation, Measure-theoretic independence, Borel-Cantelli Lemmas, Modes of probabilistic convergence, Weak and strong laws of large numbers, and Central limit theorems.

Students wishing to move on to the next level of sophistication and mathematical maturity needed for study in fields such as stochastic processes, statistics or advanced applications will find this course useful. Prerequisite: Stat-201 (MS level course in probability).

Textbooks: Karr, A. (1993). *Probability*. Springer, New York.

Supplemental Texts: Chung, K. (1974). *A Course in Probability Theory*. Academic Press, Orlando. Billingsley, P. (1990). *Probability and Measure*, 2nd Edition. Wiley, New York.

Stat 262. Nonparametric Inference. Thursday, 06:10pm-08:40pm
Instructor: Dr. S. Kundu.

This course will discuss inferential methods when the form of the underlying distribution is not specified or is only partially specified. These methods are robust as they do not rely on strong distributional assumptions. Topics to be covered in this course include: U-statistics, rank tests, locally most powerful rank tests, one and two-sample tests, asymptotic distribution theory, asymptotic relative efficiency, nonparametric point estimates and confidence intervals, goodness of fit tests. If time permits some advanced topics like Bootstrap, Nonparametric Density estimation, Nonparametric Regression will be covered.

Prerequisite: Stat 201-2 or permission of instructor.

Statistics 263. Advanced Statistical Theory I. Thursday, 6:10pm-8:40pm.
Instructor: Dr. T. Nayak.

This is an advanced course on principles and theory of statistical inference. Topics include: sufficiency, ancillarity, completeness, unbiased estimation, Cramer-Rao inequality, Bayesian estimation, admissibility, hypotheses testing.

Prerequisite: Stat 201-2 or permission of instructor.

Statistics 287. Modern Theory of Survey Sampling. Wedensday, 6:10pm-8:40pm.
Instructor: Dr. P. Chandhok

The main objectives of the course are to provide a rigorous treatment of sampling theory and its applications. With this background the student can modify the existing theory, develop new theory, and better understand its applications. Graduate students from quantitative fields such as Statistics, Mathematics, Economics, Finance and Engineering as well as professionals working in government and private-sector companies, with an interest in survey sampling will benefit from this course. The prerequisites for the class are Statistics 91 (Principles of Statistical Methods) or equivalent and Math 32 (Single-Variable Calculus) or equivalent.

This course will introduce the following topics: simple random sampling with and without replacement, systematic sampling, unequal probability sampling with and without replacement, ratio estimation, difference estimation and regression estimation.

Stat 289: Statistical Genetics. Monday, 6:10-08:40pm
Instructor: Dr. Z. Li.

There are three objectives of this course: 1) to provide an introduction of quantitative genetics for students without any genetics background; 2) to give a rigorous statistical treatment of some genetic problems; 3) to introduce current research topics in the area of statistical methods for genetic analysis.

Topics include: Allele frequency, Hardy-Weinberg equilibrium, and linkage equilibrium; Genetic variance and correlation; Parametric linkage analysis; Non-parametric linkage analysis; Recurrence-risk ratio method; The transmission/disequilibrium test (TDT); Family-based case control vs. unrelated case-control designs; Multiple point linkage analysis; Interval mapping; Haplotype-based association analysis.

Students' Corner

Hello fellow students! This month I would like to bring to your attention a mathematical/probability curiosity that I find fascinating. Strictly speaking, it might not count as statistics *per se*, but I think that students of statistics who know something about probability should find it interesting to consider.

Imagine a casino that has two tables at which you can gamble. The first table has a game, let us call it **Game A**, that has a simple coin flip – heads you win, tails you lose. This coin is weighted on one side, so that you have a probability *p* of winning, where *p* is some number between 0 and 1, and *p* is not necessarily equal to 0.5. (We're going to set *p* to a number less than 0.5, so the game is unfortunately weighted *against you*!) Below is a small MATLAB function modeling this coin toss.

```
function [ Capital ] = GameA(Capital,Bet,p)

% Generate random number in [0,1]
coinFlip  = rand(1);

% Check whether we won or lost this bet.
if ( coinFlip < p )
    Capital = Capital + Bet; % Won, increase Capital.
else
    Capital = Capital - Bet; % Lost, decrease Capital.
end
```

This simple MATLAB function takes as input the following three arguments:

- **Capital** - current amount of money that you have in your gambling account
- **Bet** - the amount of money you are betting
- **p** - the probability of winning the coin toss

and returns the amount of money that you now have left in your account after resolving the gamble. We'll assume that **Capital** and **Bet** are always integer values. Save this code in a file named **GameA.m**, in a folder that is in your MATLABPATH; if the folder isn't in your MATLABPATH, you can use the ADDPATH command to make it so.

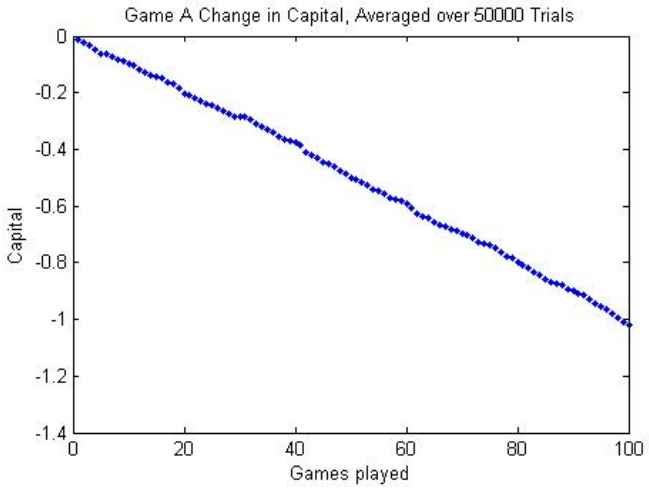
If *p* were less than 0.5 and you played Game A over and over again, you'd be guaranteed to go bankrupt in the long run; Game A would be a "losing game." The following MATLAB code simulates betting \$1 on Game A for 100 bets in a row, with *p* set to 0.5-ϵ, where ϵ=0.005. 50,000 trials of 100 sequential bets on Game A are simulated, and then averaged.

```
% TestA
T    = 50000;      % Number of trials
G    = 100;        % Number of bets made per trial.
Bet  = 1;          % Bet $1 each time.
e    = 0.005;      % Epsilon
p    = 0.5 - e;    % Game A probability of winning.
RecA = zeros(G,1); % Record of Capital fluctuations.

% Run simulations
for t=1:T;       % Loop over T trials
    Capital = 0; % Net funds for betting.
    for g=1:G;   % Loop over G games per trial
        Capital = GameA(Capital,Bet,p); % Play Game A, update
Capital
        RecA(g) = RecA(g) + Capital; % Keep cumulative record of
Capital
    end
end

% Plot results
plot(RecA/T,'b.','LineWidth',2)
title(sprintf('Game A Change in Capital, Averaged over %d
Trials',T))
xlabel('Games played')
ylabel('Capital')
```

Save the above MATLAB code into the same folder as **GameA.m**, in a file named **TestA.m**. Then invoke it in MATLAB by typing

TestA

at the MATLAB prompt. If you run this MATLAB code, you should see a plot similar to that shown to the right (results may vary a little due to the random numbers generated). This demonstrates that in the long run, you'll go bankrupt playing Game A. This is not surprising.



Now, the second table in our hypothetical casino has **Game B**, which has a rather odd coin toss game involving *two* coins, B1 and B2. In Game B, if your current capital is an integer multiple of 3, we flip coin B1; otherwise we flip coin B2. This means that on average we'll flip B2 twice as often as B1. With B1, your probability of winning is $p1 = 0.1 - \epsilon$, while with B2 it is $p2 = \frac{3}{4} - \epsilon$, where ϵ=0.005. Even though your probability of winning with coin B2 is very good, coin B1's probability of winning, *p1*, is so bad that it more than offsets *p2*. So Game B is *still* rigged against you. As a

very simple exercise, compute the overall probability of winning Game B; you should get a number less than 1, meaning that if you bet $1 the expected value of your winnings is less than $1 – a "losing game" again. Below is MATLAB code encoding Game B's scenario. Save it into a file named **GameB.m**, in the same folder where you saved **GameA.m**.

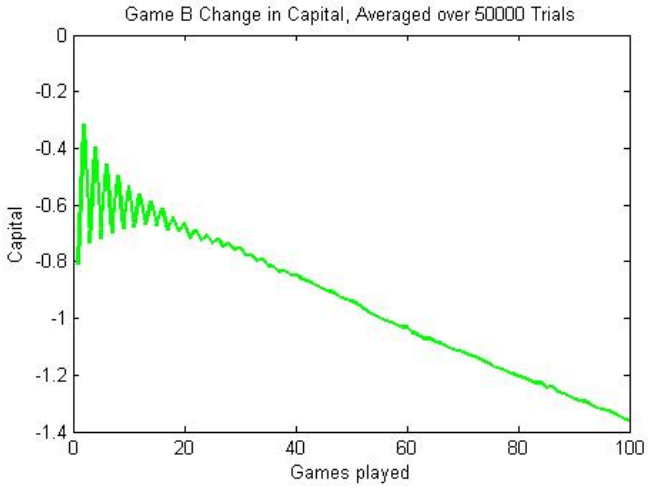
```
function [ Capital ] = GameB(Capital,Bet,M,p1,p2)

% Choose which of two coins we'll use for Game B.
if ( mod(Capital,M) == 0 )
   pp = p1; % Coin B1
else
   pp = p2; % Coin B2
end

% Generate random number in [0,1]
coinFlip  = rand(1);

% Check whether we won or lost this bet.
if ( coinFlip < pp )
   Capital = Capital + Bet; % Won, increase Capital.
else
   Capital = Capital - Bet; % Lost, decrease Capital.
end
```

The following MATLAB code simulates playing Game B. Save it into a MATLAB file named **TestB.m**, and then invoke it in MATLAB. Again, 50,000 trials of 100 sequential bets on Game B are simulated, and then averaged.

```
% TestB
T    = 50000;      % Number of trials
G    = 100;        % Number of bets made per trial.
B    = 1;          % Bet $1 each time.
e    = 0.005;      % Epsilon
M    = 3;          % Modulus base.
p1   = (1/10) - e; % Game B, Coin B1.
p2   = (3/4) - e;  % Game B, Coin B2.
RecB = zeros(G,1); % Record of Capital fluctuations.

% Run simulations
for t=1:T;       % Loop over T trials
    Capital = 0; % Net funds for betting.
    for g=1:G;   % Loop over G games per trial
        Capital = GameB(Capital,B,M,p1,p2); % Play Game B, update
Capital
        RecB(g) = RecB(g) + Capital; % Keep cumulative record of
Capital
    end
end
```

```
% Plot results
plot(RecB/T,'g-','LineWidth',2)
title(sprintf('Game B Change in Capital, Averaged over %d
Trials',T))
xlabel('Games played')
ylabel('Capital')
```

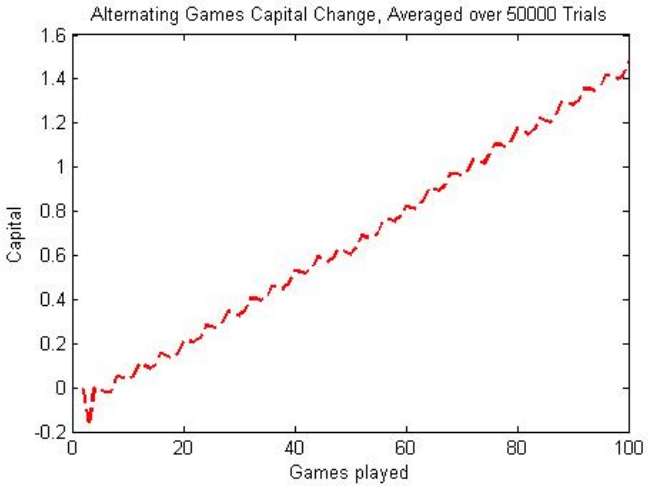
If you run this MATLAB code, you should see a plot similar to that shown to the right. As with Game A, in the long run you'll lose money in the long run with Game B. Again, this is not surprising. (By the way, what do you make of those oscillations at the beginning of the plot, where the number of games played is less than 20?)



But a very interesting thing happens if you now alternate between two bets on Game A and two bets on Game B: all of the sudden you start winning money! To demonstrate this save the MATLAB code shown below in a file named **TestAB.m**, and then invoke it in MATLAB. This code simulates switching between two gambles on Game A and then two on Game B, 100 times in a row. Again, 50,000 simulations are run and then averaged.

```
% TestAB
T     = 50000;      % Number of trials
G     = 100;        % Number of bets made per trial.
B     = 1;          % Bet $1 each time.
e     = 0.005;      % Epsilon
p     = 0.5 - e;    % Game A probability of winning.
M     = 3;          % Modulus base.
p1    = (1/10) - e; % Game B, Coin B1.
p2    = (3/4) - e;  % Game B, Coin B2.
RecAB = zeros(G,1); % Record of fund fluctuations.

count = 0; % Initialize counter.
game  = 0; % IF game==0, Game A; IF game==1, Game B

% Run simulations
for t=1:T;       % Loop over T trials
    Capital = 0; % Net funds for betting.
    for g=1:G;   % Loop over G games per trial
        count = count + 1;

        % Check whether we're playing game A.
        if ( game == 0 )
```

```
            Capital = GameA(Capital,Bet,p); % Play Game A, update
Capital

            % Check whether to switch to game B now.
            if ( count == 2 )
                count = 0;
                game  = 1;
            end

        % Else we're playing game B.
        else

            Capital = GameB(Capital,B,M,p1,p2); % Play Game B,
update Capital

            % Check whether to switch to game A now.
            if ( count == 2 )
                count = 0;
                game  = 0;
            end
        end; % IF game

        RecAB(g) = RecAB(g) + Capital; % Keep cumulative record of
Capital
    end; % FOR g
end; % FOR t

% Plot results
plot(RecAB/T,'r--','LineWidth',2)
title(sprintf('Alternating Games Capital Change, Averaged over %d
Trials',T))
xlabel('Games played')
ylabel('Capital')
```



This MATLAB code should produce a plot similar to that shown on the right. As you can see, if you alternate between two bets on Game A and two on Game B, you now start winning money and your capital steadily increases! This paradoxical effect is known as *Parrondo's Paradox*, or *Parrondo's Games*.

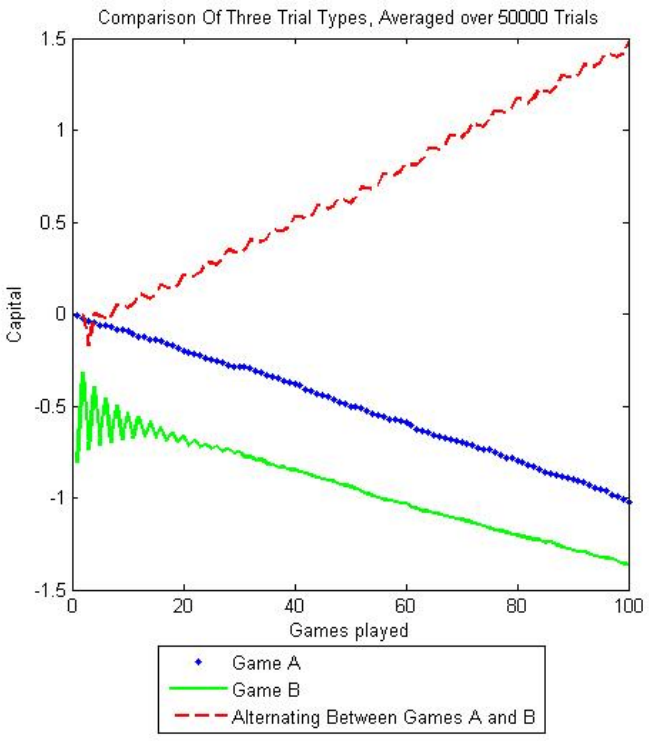
Do you not find this rather surprising? The effect still holds if you randomly switch between the two games, rather than regularly alternate between them. Modify the MATLAB code to demonstrate this! Can you then create a plot that replicates Figure 1b in Harmer and Abbott's 1999 article in *Nature*? (By the way, this installment of the *Students' Corner* is heavily

based on Harmer and Abbott's article. Please try reading it; it is less than a page long, and highly readable!) Of course, Game B is rather contrived, since it is conditional on whether your current capital is divisible by 3. In a sense, Game B "knows" something about how much capital you currently have, which is kind of strange.

To combine the three previous graphs into one MATLAB plot, try the following MATLAB commands:

```
axis([0 100 -1.5 1.5])
plot(RecA/T,'b.','LineWidth',2)
hold on
plot(RecB/T,'g-','LineWidth',2)
plot(RecAB/T,'r--','LineWidth',2)
hold off
title(sprintf('Comparison Of Three Trial Types, Averaged over %d Trials',T))
xlabel('Games played')
ylabel('Capital')
legend('Game A','Game B',...
    'Alternating Between Games A and B','Location','SouthOutside')
```

An example of running this snippet of MATLAB code is shown to the right.



Try varying the values of *p*, *p1*, and *p2*. Does the paradoxical effect hold for all values of *p*, *p1*, and *p2*? Does the value of the modulus base M matter? What happens if you increase ϵ? Does ϵ have to have the same value for both Games A and B? Suppose instead of betting exactly $1 each time, you instead bet a *constant fraction* of your current holdings, e.g. one one-hundredth of your current capital. What happens if you modify **TestA.m** to model this? What about **TestB.m**?

Now consider this thought experiment. Suppose a husband and wife team enter our hypothetical casino, and the man plays Game A while the woman plays Game B. Since they're married, they share the same gambling account, and Game B is conditional on whether the amount in that joint account is divisible by 3. Would the husband and wife working as a team make money? Now suppose the couple have a falling out, and they decide to hold separate gambling accounts, but they continue to play only at their respective tables, the husband at Game A and the wife at Game B. Would it make a difference that they now have separate gambling accounts?

So why does Parrondo's Paradox work? See Chapter 11 in Julian Havil's book, *Nonplussed! Mathematical Proof of Implausible Ideas*, for a highly detailed explanation of how this phenomenon arises. I must admit that I haven't yet had the patience to wade through the algebra – it's a little dense – but I thought I'd alert you to Dr. Havil's explanation, if you wanted to try reading it!

The big question that you must be wondering at this point is, *can I use this to make money in the stock market*? It has been argued that you can't use Parrondo's Paradox to make money in the stock market (Iyengar and Kohli, 2004). However, a New York Times article published in 2000 reported that Dr. Sergei Maslov, a physicist at Brookhaven National Laboratory, showed that it might indeed be possible. I emailed Dr. Maslov asking him for further information on this finding, and he directed me to the bottom of the 5th page of a 1998 paper he published in the International Journal of Theoretical and Applied Finance. You can find that paper at this URL: http://xxx.lanl.gov/abs/cond-mat/9801240 . Perhaps you can figure out how to use Parrondo's Paradox to make some money, and then you can pay your tuition!

Harmer and Abbott conclude their 1999 *Nature* article with the following intriguing conjecture:

> Game theory is linked to various disciplines such as economics and social dynamics, so the development of parrondian-like strategies may be useful, for example for modelling cases in which declining birth and death processes combine in a beneficial way.

That's all for this month. If you have any feedback on this column or ideas for future topics, please email me at jmm97@georgetown.edu. As always, your thoughts will be greatly appreciated.

Joe Maisog

Georgetown University / Medical Numerics

References:

Harmer GP and Abbot D, Losing strategies can win by Parrondo's paradox, Nature 402(864):864, 1999.

Havil J, *Nonplussed! Mathematical Proof of Implausible Ideas*, Princeton, NJ:Princeton University Press, 2000, Chapter 11 (pp. 115-126).

Iyengar R and Kohli R, Why Parrondo's Paradox Is Irrelevant for Utility Theory, Stock Buying, and the Emergence of Life, Complexity 9(1):23-27, 2004.

Maslov S and Zhang YC, Optimal investment strategy for risky assets, International Journal of Theoretical and Applied Finance 1(3):377-387 (1998).

Paradox in Game Theory: Losing Strategy That Wins, New York Times (Science Times section), January 25, 2000.

See also:
Harmer et al., The Paradox of Parrondo's Games, Proceedings: Mathematical, Physical and Engineering Sciences, Vol. 456, No. 1994. (Feb. 8, 2000), pp. 247-259.

Harmer and Abbott, Parrondo's Paradox, Statistical Science, Vol. 14, No. 2. (May, 1999), pp. 206-213. See especially section 1.1, which makes an interesting connection between a theoretical physical device called a *Brownian Ratchet* (which seems sort of like a free energy device!) and Parrondo's Paradox.

For an interactive Java simulation, see the web page http://www.cut-the-knot.org/ctk/Parrondo.shtml (A. Bogomolny, *Parrondo Paradox*, from *Interactive Mathematics Miscellany and Puzzles*).

From reading the description on Amazon.com, I gather that Richard Armstrong's 2006 novel *God Doesn't Shoot Craps: A Divine Comedy* is about a fictional gambler who uses Parrondo's Paradox to make money. I think I will buy this book and read it!

Employment

As a service to local statisticians, *WSS News* provides notification of employment opportunities and description of those seeking employment here in the Washington, DC, area. Readers are encouraged to take advantage of this feature of the newsletter. The deadline for inserting notices is five (5) weeks before the publication date. Those interested should email or call Anne Peterson, at apeterson@insightpolicyresearch.com or (703) 373-6645.

CLINICAL TRIAL BIOSTATISTICIANS
M.S. and Ph.D. Level Positions

With an opportunity for substantial leadership responsibility in studies of international public health import.

The Biostatistics Center of The George Washington University, founded in 1972, is a leader in the statistical coordination of clinical trials conducted by the National Institutes of Health. We enjoy over $45 million per year of NIH research funding for major studies in cardiovascular disease, diabetes, maternal/fetal medicine, osteoporosis, urology, and the genetic basis for various diseases. The center has a staff of over 100 with 27 biostatisticians/epidemiologists, including 10 faculty. We are recruiting M.S. and Ph.D. level staff to participate in these and future studies. Please visit our web site (below).

Master's Level Research Positions: These positions require a Master's in Biostatistics or Statistics and 1-5 years experience in analysis, supervision of data management and study design for biomedical applications. Good written and oral communication skills, and detailed knowledge of SAS required. Send CV to address below.

Assistant to Full Research Professorial Positions are available immediately to serve as Co-Investigator or Principal Investigator (Project Director) and to provide statistical direction of the design, conduct and analysis of studies and the conduct of methodologic research to meet the projects needs. We are seeking individuals who want to join a highly competent team of academic biostatisticians and epidemiologists; who desire to contribute to the design and analysis of major medical studies, seek substantive scientific and statistical responsibility, enjoy interacting with medical investigators; take pride contributing to the publication of major papers in leading medical journals, and desire to make an impact on the public health. Our faculty also participate in graduate programs in biostatistics, epidemiology and statistics which afford opportunities for teaching at the graduate level. The research projects also provide an environment rich in methodological problems, with opportunities for collaboration with research active Center faculty and graduate students.

Minimum Position Requirements: Doctorate in Biostatistics, Statistics or Epidemiology, or alternatively an M.D. or Ph.D. in Biological Science, Physical Science or Computer Science with a Masters in Biostatistics or Statistics, 1-5 years' experience with clinical trials, especially study design and statistical analysis of study results using SAS, excellent oral and written English

communication skills, and supervisory experience.

Application Procedures: Applicants must send a Curriculum Vitae and three letters of reference; a letter to include a synopsis of their role in collaborative medical research that has led to medical scientific presentation or publication and a statement of career purpose indicating their career goals and how this position can help you achieve those goals; and applicants for Assistant Research Professor positions must send an Official Transcript of graduate coursework leading to the doctoral degree to: Sarah Fowler, Research Professor and Director, The George Washington University Biostatistics Center, 6110 Executive Blvd., Suite 750, Rockville, MD 20852.

HTTP://WWW.BSC.GWU.EDU

Review of applications is ongoing until the positions are filled. Rank/position title and salary commensurate with experience and qualifications. Tuition benefits for employees (including Ph.D. in Statistics, Biostatistics and Epidemiology) and for spouse and dependent children.

All research and regular faculty at the rank of Assistant Professor in Biostatistics or Statistics may apply for the Samuel W. Greenhouse Biostatistics Research Enhancement Award. For a period of 1 year, the award will provide 20% effort for methodological research and a discretionary fund to support professional activities, travel to professional meetings, supplies and equipment. Applicants for the research faculty position may also apply for the Greenhouse Award while their faculty application is being considered. For complete information including Award Application Materials Requirements, please visit our website at: www.bsc.gwu.edu.

The George Washington University is an Equal Opportunity/Affirmative Action employer

Survey Sampling Statistician

WESTAT: AN EMPLOYEE-OWNED RESEARCH CORPORATION

Westat is an employee-owned corporation headquartered in the suburbs of Washington, DC (Rockville, Maryland). We provide statistical consulting and survey research to the agencies of the U.S. Government and to a broad range of business and institutional clients. With a strong technical and managerial staff and a long record of quality research, Westat has become one of the leading survey research and statistical consulting organizations in the United States.

Our company was founded in 1961 by three statisticians. The current staff of more than 1,800 includes over 60 statisticians, as well as research, technical, and administrative staff. In addition, our professional staff is supported by data collection and processing personnel situated locally and in field sites around the country. The work atmosphere is open, progressive, and highly conducive to professional growth.

Our statistical efforts continue to expand in areas such as the environment, energy, health, education, and human resources. Westat statisticians are actively involved in teaching graduate-level courses in statistical methods and survey methodology in collaborative arrangements with area colleges and

universities.

We are currently recruiting for the following statistical position:

Survey Sampling Statistician (Job Code WSS/DRM/7001)

Three or more years of relevant experience in sample design and selection, frames development, weighting, imputation, and variance estimation. Must have a master's or doctoral degree in statistics and have excellent writing skills. Coursework in sample survey design is highly desirable.

Westat offers excellent growth opportunities and an outstanding benefits package including life and health insurance, an Employee Stock Ownership Plan (ESOP), a 401(k) plan, flexible spending accounts, professional development, and tuition assistance. For immediate consideration, please send your cover letter, indicating the Westat Job Code, and resume by one of the following methods to:

Job Code is REQUIRED to apply.
Westat • Attn: Resume System • 1650 Research Boulevard • Rockville, MD 20850-3195
Email: resume@westat.com • FAX: (888) 201-1452
Equal Opportunity Employer. www.westat.com

Chief Mathematical Statistician
Social Security Administration

The Social Security Administration's Office of Research, Evaluation, and Statistics is seeking a Chief Mathematical Statistician. In addition to full-time, SSA employment, SSA will consider a detail from other Federal Agencies or an Intergovernmental Personnel Act (IPA) agreement with an academic institution. The Chief Mathematical Statistician reports to the Associate Commissioner and serves as the senior consultant and expert on mathematical and statistical matters. The incumbent recommends and implements statistical policies, assures that appropriate methodologies and practices are used, and serves as ORES' spokesperson on statistical matters.

The Chief Mathematical Statistician must have in-depth, expert knowledge of statistical theory and methods and skill applying this knowledge to highly complex, technical projects, such as:

- Providing advice and serving as lead consultant on the design and use of sample surveys, including sample design, estimation, data analysis, reliability measures, and survey costs/timing.
- Serving as ORES' lead expert and advisor on statistical disclosure avoidance methods, and working with staff to develop, design, and implement current methods.
- Leading the design and implementation of procedures to create public use microdata samples from administrative data.
- Linking data from two independent sources, such as surveys and administrative records, to develop improved estimates.
- Representing ORES at meetings, seminars, and federal statistical working groups.

For more information please contact: Manuel de la Puente, Associate Commissioner, Office of Research, Evaluation, and Statistics at 202/358-6020 or at: manuel.de.la.puente@ssa.gov

Note from the WSS NEWS Editor

Items for publication in the Summer issue of the WSS NEWS will be accepted until July 7, 2008. E-mail items to Michael Feil at michael.feil@usda.gov.

P.O. Box 752
Suitland, MD 20752

NON-PROFIT
ORG.
U.S. POSTAGE
PAID

Washington
Statistical
Society

Washington Statistical Society Annual Dinner

Wednesday June 25, 2008, 6:30 PM
Cocktails 6:00 PM

Meiwah Restaurant, 4457 Willard Ave, Chevy Chase, MD
(Metro: Friendship Heights on the Red Line)

Keynote Presentation

by

Thomas Lumley, University of Washington

Open Source Statistical Software: Why? When? Where?

- Made Possible by Funding from RTI International -

Cash Bar

Variety of Dishes Including:

Crispy Fried Shredded Beef

And

Chicken with Black Bean Sauce

And

Shanghai Bokchoi with Fresh Mushrom

Price:……. $45.00 per person

Please register at: https://www.123signup.com/event?id=tmkbh

Or call/email: Yves Thibaudeau, (301) 763-1706, Yves.Thibaudeau@census.gov