Calculated Bets: Computers, Gambling, and Mathematical Modeling to Win

Steven Skiena

Department of Computer Science
State University of New York
Stony Brook, NY 11794–4400
http://www.cs.sunysb.edu/∼skiena

28th Atlantic Provinces Council APICS
October 16, 2004

Program Trading in Jai-Alai

Investing in the financial markets is often likened to gambling.

This is story of how I used computer simulations and mathematical modeling techniques to bet on jai-alai.

I tell it to introduce several things which interest me:

- *The Joys of Jai-Alai* Jai-alai is a spectator sport and gambling venue which is under-appreciated by the public.
- *The Power of Mathematical Modeling* Mathematical models govern our economy and help forecast our weather, but the man on the street has little understanding of what they are and how they work.
- *The Craft of Computer Programming* Most people do not understand how computer programs written, and the elegance and beauty behind all interesting software.

- *The Mathematics of Money* Gambling and mathematics have a long and interesting history together. My jai-alai system functions very much as a miniature stock trading system.
- *The Aesthetics of Data* – Data analysis is both an art and a science. The right data representation lets us hear what the numbers are trying to tell.

Finally, this is the story of a professor who bets real money to test if his system works.

This is joint work with my students Roger Mailler, Meena Nagarajan, Dario Vlah, and San-Yiu Cheng.

Early History

I first learned to program a computer as a high school student back in 1977, when I developed a simple little model to predict the results of NFL football games.

Simply averaging points for/against and giving three points for home field will predict the winner of 65% of all NFL games.

This was a big enough deal back then for me to become a newspaper columnist. . .

Student uses computers to predict football winners

By JEFF LEEBAW
Home News staff writer

EAST BRUNSWICK — A 16-year-old East Brunswick High School student has found a way to combine an interest in football with a fascination for computers.

Steven Skiena says he can determine, with a high degree of accuracy, the outcome of professional football games by feeding a computer pertinent information about competing teams.

"The winners will almost always be correct," said the high school junior who lives at 5 Currier Road off Dunhams Corner Road. "I had an 86 per cent accuracy rate when I started predicting at the end of last season."

He does it by feeding the computer a myriad of statistics that include team records, points scored and allowed, average yards gained and allowed during a game, a breakdown of the yards gained and allowed into rushing and passing categories, performances at home and on the road, and more.

The information is gathered from weekly compilations of football statistics and standings. Skiena puts the facts on index cards and then types them into one of the six computer terminals at the high school or a terminal at The Library where he works part-time after school.

"I get a winning team, a decimal score for each team and a point spread," said the teen-ager who completed a computer programming course last year at the high school.

His first attempt at picking winners involved a Monday night game between the Oakland Raiders, the eventual Super Bowl victors, and the Cincinnati Bengals.

It was a difficult game to analyze because Cincinnati was fighting for a playoff berth while Oakland had already clinched a spot in the post season competition.

"Nobody knew whether Oakland would be giving 100 per cent," Skiena said. "But my calculations indicated they would win by 24-20. The final score was 35-20. They went all out."

Skiena said he went on to pick 12 of the 14 winners the following week and accurately predicted Oakland would defeat Minnesota in the Super Bowl.

The National Football League's 1976 Record Book, which breaks down last year's statistics for each of the league's 28 teams, will supply most of Skiena's information for the first few weeks of the 1977 season. He will also use statistics from the final two exhibition games played this year by each of the teams.

Skiena wrote a computer program based on 17 statistical variables that might come into play during a football game.

STEVEN SKIENA
...to test his accuracy

Predictions published

Steven Skiena will get a chance to display his skill as a pro football prognosticator each Sunday in The Home News.

The youngster's weekly selections will be an "added ingredient to our football coverage," according to Home News Executive Editor Robert E. Rhodes.

"I think it's interesting enough for us to give it a

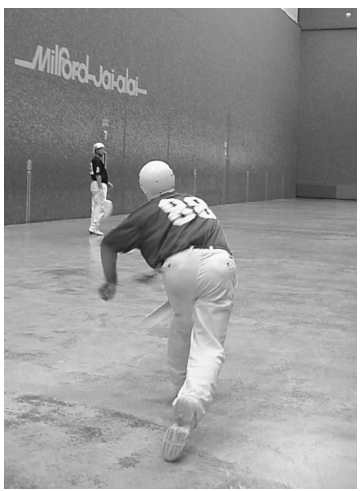
Lessons from Football Prediction

1. Simple mathematical models have real predictive power
2. Simple mathematical models do not have predictive power when money is on the line

I also got interested in jai-alai during high school, after being exposed to the game on a family vacation to Florida.

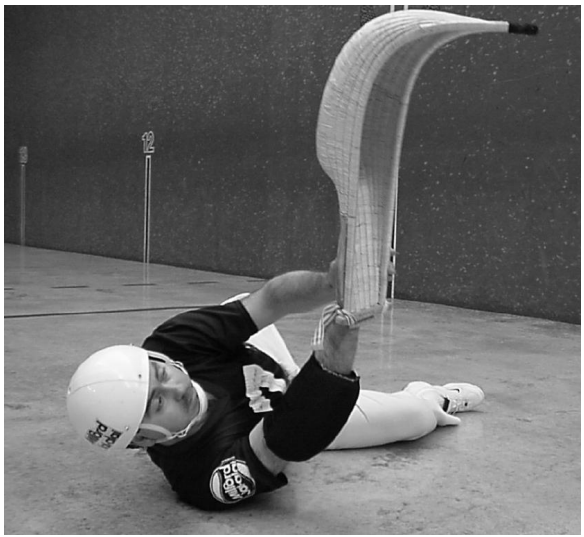
What is Jai-Alai?

Jai-alai is a sport of Basque origin also known as *Pelota Vasca* or *Cesta Punta*.



It is a variation on hand ball, where the rubber-and-goatskin *pelota* is caught/thrown with an large basket or *cesta*, and bounced off a granite or concrete wall.

The game is extremely fast – pelotas have been clocked at over 180 miles-per-hour, or twice as fast as a major league fastball.



Jai-alai betting is legal in Connecticut, Florida, and Rhode Island, as well as the Basque country of Spain and France.

Why Jai-Alai?

The scoring system in use in the United States is different than that of the Basques, and particularly interesting to a mathematician or computer scientist.

As a parimutuel sport, jai-alai has evolved to permit more than two players in a match.

The Spectacular Seven Scoring System

Eight players or teams play in each match, arranged in a first-in, first-out (FIFO) queue, with two teams playing each point. The loser of the point is added to back of the queue, with the winner staying on to play the team at the top. Play continues until one team totals (typically) seven points.

To reduce the disadvantage of late post positions, the *Spectacular Seven* scoring system increases the reward for each score after the seventh physical point from 1 to 2 points.

Spectacular Seven Example

Queue						Point	Score							
1	2	3	4	5	6	winner	1	2	3	4	5	6	7	8
3	4	5	6	7	8	1-beats-2	1	0	0	0	0	0	0	0
4	5	6	7	8	2	1-beats-3	2	0	0	0	0	0	0	0
5	6	7	8	2	3	1-beats-4	3	0	0	0	0	0	0	0
6	7	8	2	3	4	5-beats-1	3	0	0	0	1	0	0	0
7	8	2	3	4	1	6-beats-5	3	0	0	0	1	1	0	0
8	2	3	4	1	5	7-beats-6	3	0	0	0	1	1	1	0
2	3	4	1	5	6	8-beats-7	3	0	0	0	1	1	1	1
3	4	1	5	6	7	8-beats-2	3	0	0	0	1	1	1	3
4	1	5	6	7	2	3-beats-8	3	0	2	0	1	1	1	3
1	5	6	7	2	8	3-beats-4	3	0	4	0	1	1	1	3
5	6	7	2	8	4	1-beats-3	5	0	4	0	1	1	1	3
6	7	2	8	4	3	5-beats-1	5	0	4	0	3	1	1	3
7	2	8	4	3	1	5-beats-6	5	0	4	0	5	1	1	3
2	8	4	3	1	6	5-beats-7	5	0	4	0	7	1	1	3
-	-	-	-	-	-	5-1-3	*5*	0	*4*	0	*7*	1	1	3

Note that 1 “won” as many points as 5, but not when they counted for as much.

Breaking Ties

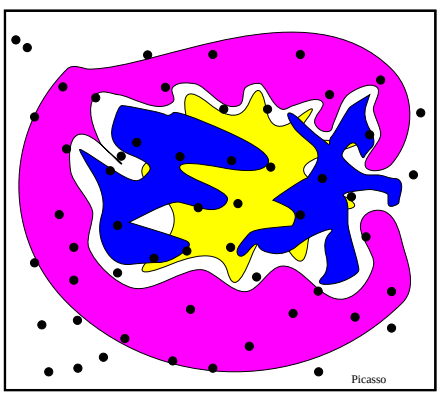
Various tie-breaking strategies are used to determine the place and show positions.

Queue						Point	Score							
1	2	3	4	5	6	winner	1	2	3	4	5	6	7	8
3	4	5	6	7	8	1-beats-2	1	0	0	0	0	0	0	0
4	5	6	7	8	2	1-beats-3	2	0	0	0	0	0	0	0
5	6	7	8	2	3	1-beats-4	3	0	0	0	0	0	0	0
6	7	8	2	3	4	5-beats-1	3	0	0	0	1	0	0	0
7	8	2	3	4	1	6-beats-5	3	0	0	0	1	1	0	0
8	2	3	4	1	5	7-beats-6	3	0	0	0	1	1	1	0
2	3	4	1	5	6	7-beats-8	3	0	0	0	1	1	2	0
3	4	1	5	6	8	2-beats-7	3	2	0	0	1	1	2	0
4	1	5	6	8	7	2-beats-3	3	4	0	0	1	1	2	0
1	5	6	8	7	3	4-beats-2	3	4	0	2	1	1	2	0
5	6	8	7	3	2	1-beats-4	5	4	0	2	1	1	2	0
6	8	7	3	2	4	1-beats-5	*7*	*4*	0	2	1	1	2	0
-	-	-	-	-	-	7-beats-4	*7*	*4*	0	2	1	1	*4*	0
-	-	-	-	-	-	1-2-7	*7*	*4*	0	2	1	1	*4*	0

Since from 2 to 7 teams can all tie, the tie-breaking rules are fairly complicated.

Monte Carlo Simulation

Monte Carlo techniques use many random experiments to approximately measure a real quantity, as in Monte Carlo integration.



The biases inherent in the Spectacular Seven Scoring System can be identified by performing a Monte Carlo simulation.

Simulating Jai-Alai

To measure scoring system bias, we can select the winner of each point by a random number generator.

For each game, we must maintain the current score and queues, and tabulate the results:

Initialize the current players to 1 and 2.
Initialize the queue to players $\{3, 4, 5, 6, 7, 8\}$.
Initialize the point total for each player to zero.
So long as the current winner has less than 7 points:
    Play simulated point between two current players.
    Add one (or if $>$ 7th point, two) to the point
        total of the winner.
    Put the loser at the end of the queue.
    Get the next player off the front of the queue.
End.
Identify current point winner as match winner.

Simulation Results

Pos	% Wins	% Places	% Shows
1	16.27%	17.93%	15.18%
2	16.30%	17.86%	15.27%
3	13.91%	16.57%	14.66%
4	12.45%	13.32%	13.79%
5	10.20%	10.83%	12.92%
6	10.27%	7.82%	11.07%
7	8.86%	8.21%	8.87%
8	11.75%	7.45%	8.24%

- 1 and 2 have a substantial advantage over the rest of the field. Either of the initial players are almost twice as likely to succeed than position 7.
- Doubling the value of each point after the seventh point in the match improves player 8's chances to win. But the real beneficiaries of point doubling are 1/2.

Actual Results

Four years of win results from Berenson's Jai-alai:

Pos	1983	1984	1985	1986	total	%wins
1	437	387	451	475	1750	14.1
2	459	403	465	486	1813	14.6
3	380	403	374	435	1592	12.8
4	351	345	368	361	1425	11.5
5	371	370	357	389	1487	12.0
6	329	414	396	402	1541	12.4
7	308	371	348	343	1370	11.1
8	357	366	351	331	1405	11.3
totals	2992	3059	3110	3222	12383	100.0

As predicted 1 and 2 do best, 7 the worst.

The bias predicted by our simulations is reduced by the matchmakers who schedule the games.

Betting on Jai-Alai

Jai-alai has been called “a lottery with seats”.
Each fronton supports a variety of different types of bets:

- *Win*, *Place*, *Show* – You can bet your team finishes either first, second, or third.
- *Quiniela* – You pick two teams to finish first and second. The order of finish is irrelevant

  Personally, I find the quiniela bet to be the most exciting choice for the spectator.
- *Exacta* – Pick two teams to finish first and second in that given order.

- *Trifecta* – Pick the three teams to finish first, second and third, in that exact order. There are $8 \times 7 \times 6 = 336$ possible trifecta bets. Trifectas are the riskiest conventional bet, but pays the highest returns.

Our system mostly bets trifectas.

Lesson: The greatest inefficiencies in a market reside in the most complex derivatives.

Trifecta Results

Certain trifectas occur 1,000 times more often than others!

- The best trifectas are 1-3-2, 1-4-2, 1-5-2, 4-1-3, and their symmetrical variants, 2-3-1, 2-4-1, 2-5-1, and 4-2-3.
- Several trifectas are *unbelievably* terrible bets, such as 5-7-8, 5-8-7, 6-7-8, and 6-8-7. In 29,096 games at Dania jai-alai, these four trifectas (and only these four) never happened.
- There is a strong bias against neighboring players.

Getting and Parsing Data

These results are not enough to bet successfully, because such biases are basically captured in the daily odds.

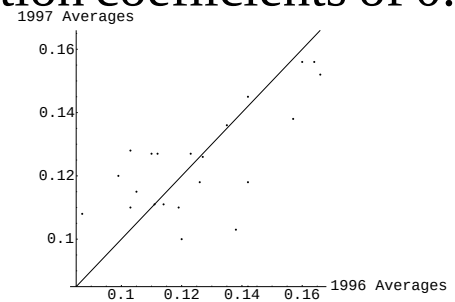
We need to analyze statistical data to measure player skills and estimate payoffs.

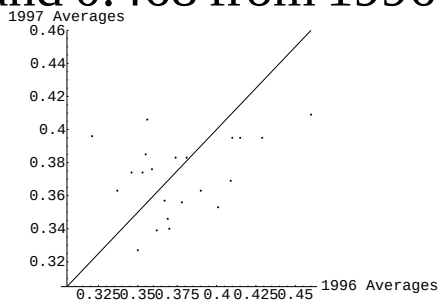
Each night our program gets the latest data from the WWW, parses it using Perl programs, and adds it to our database.

We now have collected more than four years of data for analysis.

Are There Good Players and Bad Players?

How consistent are jai-alai players from year to year? Milford jai-alai player win and in-the-money percentages had correlation coefficients of 0.492 and 0.468 from 1996 to 1997.





Compare to correlations of baseball batting averages (0.587 AL / 0.517 NL) and pitching ERAs (0.386 AL / 0.503 NL). Thus we *can* use history to evaluate player skills in a meaningful way.

How Can We Rate Players?

By combining the players record in each starting position with data from our simulation, we can estimate what fraction of points each player wins.

Pos	Winner	Placer	Shower	Other
1	78.77%	63.52%	51.76%	27.26%
2	78.58%	63.29%	51.65%	26.73%
3	79.58%	66.16%	54.63%	27.91%
4	81.59%	67.75%	57.64%	29.45%
5	82.94%	70.32%	60.72%	30.83%
6	85.84%	70.18%	63.09%	31.80%
7	87.21%	72.05%	63.89%	31.76%
8	89.44%	70.41%	62.75%	29.96%
all	82.08%	66.88%	57.16%	29.48%

How Often Does x Beat y?

The Monte Carlo simulation needs to know how often player x wins against player y.

All we know is that x wins 54.8% of his points and y wins 49.3% of his, both against the world.

To combine these, we need a formula which generates probabilities with the following properties:

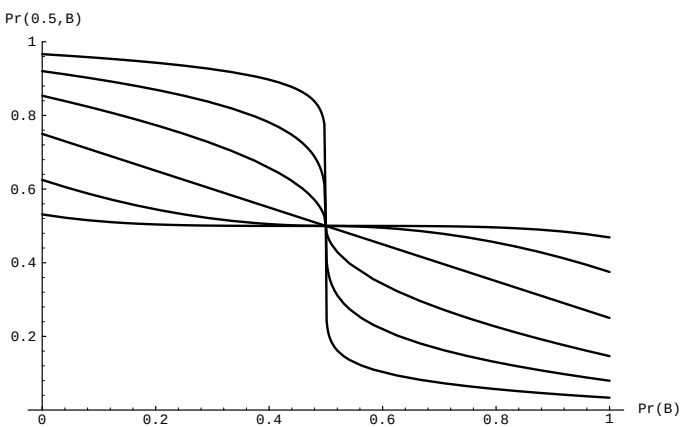
- *Integrity* – $Prob(A, B)$ must be a real probability value.
- *Symmetry* – $Prob(A, B) = 1 - Prob(B, A)$
- *Equality* – $P(A) = P(B) \rightarrow Prob(A, B) = 1/2$.
- *Generality* – We need a "fudge factor" to easily change the shape of the desired solution.

The Cesta of Skiena

The following function satisfies all of these requirements:

$$Prob(A,B) = \frac{1 + (P(A) - P(B))^{\alpha}}{2} \quad \text{if} \quad P(A) \geq P(B)$$

$$Prob(A,B) = \frac{1 - (P(B) - P(A))^{\alpha}}{2} \quad \text{if} \quad P(A) \leq P(B)$$



How Much Does Each Bet Pay?

Jai-alai frontons make their money by taking a cut of all bets, under the *parimutuel* system.

All money wagered on each type of bet is put into a pool, and divided among the winners according to how much they bet.

But first the house takes 20% of this pool.

Thus we must do 20% better than average to make money.

Frontons want you to bet, casinos want you to lose. Thus they were happy to take my bets even if my system wins.

By averaging the results of previous payoffs, we can estimate the likely payoff for every possible bet.

How Much Do We Impact the Pool?

Our analysis shows that a small number of bets have positive expected return. Should we bet a lot of money on these?

No! Our successful bets will lower the payoffs on these outcomes, eventually leading to negative expected return.

This is why real investors bet at the stock market instead of the fronton. Jai-alai pools are small enough that very modest investments flood the market.

This is why we will not get rich from jai-alai.

By estimating the total amount bet in each pool, we can predict the impact of our own wagering.

Putting the System Together

Each night, our system down loads the schedule of tomorow's matches.

For each game, we evaluate the skills of each player, and then run our Monte Carlo simulation 500,000 times.

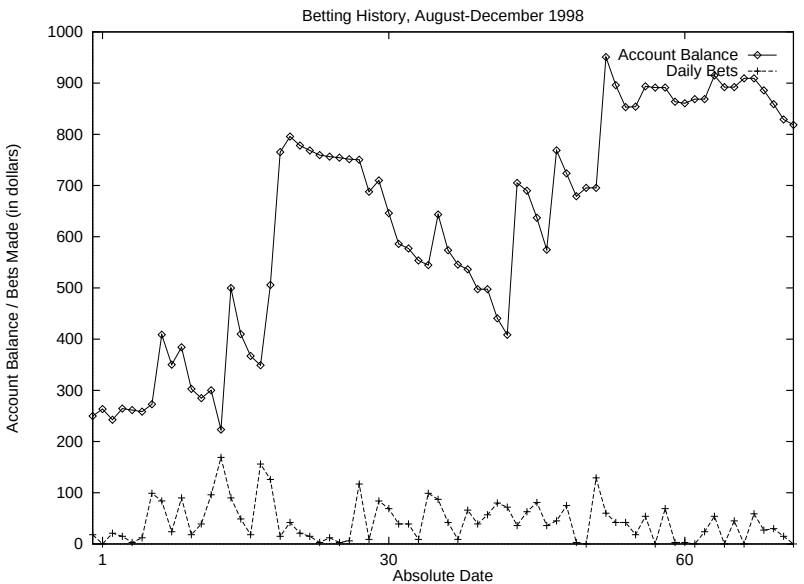
We compare the results with our payoff model to determine the most advantageous bets.

In the morning, our computer phones the bets in (using a modem) to the fronton's computer.

There is no human involvement in analysis or wagering!

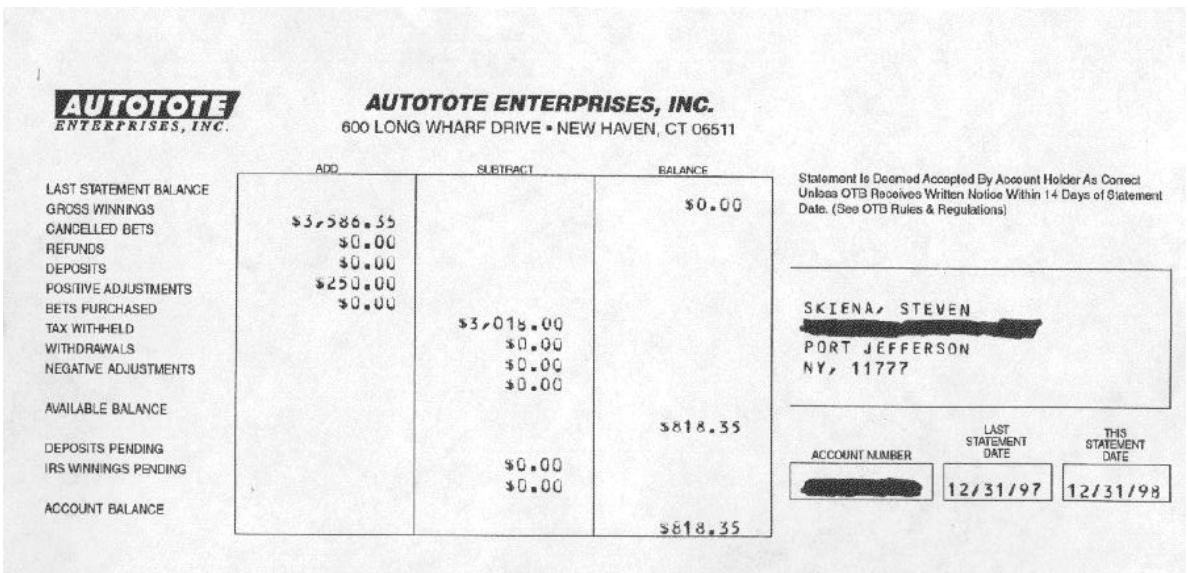
How Did We Do?

We carefully logged our bets from July to December, 1998.



We turned an initial $250 into $818.35, making a total of $3018 of bets.
This is a total capital return of 227% over the period, with a return of 18% per bet.
A revised system later increased our total to $1851.70 before breaking due to a software bug.

Proof of Our Winnings



AUTOTOTE ENTERPRISES, INC.

AUTOTOTE ENTERPRISES, INC.
600 LONG WHARF DRIVE • NEW HAVEN, CT 06511

	ADD	SUBTRACT	BALANCE
LAST STATEMENT BALANCE			$0.00
GROSS WINNINGS	$3,586.35		
CANCELLED BETS	$0.00		
REFUNDS	$0.00		
DEPOSITS	$250.00		
POSITIVE ADJUSTMENTS	$0.00		
BETS PURCHASED		$3,015.00	
TAX WITHHELD		$0.00	
WITHDRAWALS		$0.00	
NEGATIVE ADJUSTMENTS		$0.00	
AVAILABLE BALANCE			$818.35
DEPOSITS PENDING		$0.00	
IRS WINNINGS PENDING		$0.00	
ACCOUNT BALANCE			$818.35

Statement Is Deemed Accepted By Account Holder As Correct Unless OTB Receives Written Notice Within 14 Days of Statement Date. (See OTB Rules & Regulations)

SKIENA, STEVEN
[redacted]
PORT JEFFERSON
NY, 11777

ACCOUNT NUMBER	LAST STATEMENT DATE	THIS STATEMENT DATE
[redacted]	12/31/97	12/31/98

Epilogue

We recently restored the system to working order, including the Dania and Miami Frontons – however betting at these frontons might violate Federal law!

Unfortunately, Milford Jai-alai closed in December 2001 and Newport in July 2004.

All winnings have been donated to charity.

I have heard from many interesting readers of my book. . .

Lesson: Google people before you get involved with them!

Future Work

I have recently set up a Financial Engineering Laboratory at Stony Brook to study more interesting markets.
Initial projects include:

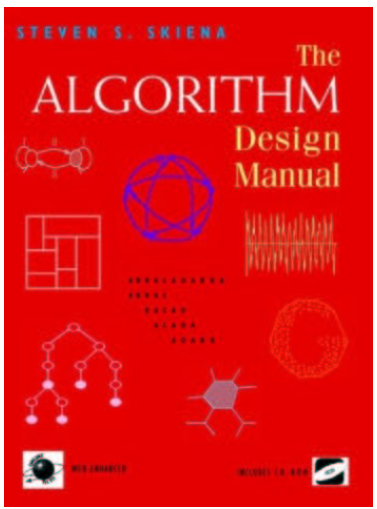
- Market timing strategies which avoid the use of sophisticated options.
- Data integration/analysis from eclectic web sources (e.g. Google, Amazon), with associated time series correlation analysis.
- Natural language analysis of news streams.

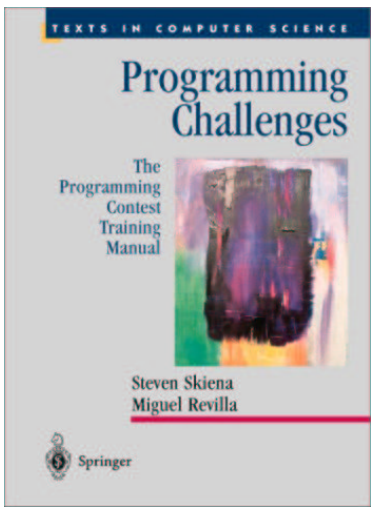
Both Monte Carlo analysis and web scraping techniques are applicable in these venues.

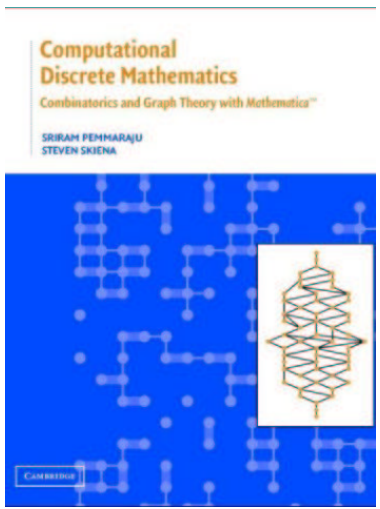
Advice to Young Computer Scientists

- Learn by doing cool projects…
- Don’t worry so much about what others have done…
- Be broad, not narrow…
- Go to a good graduate school, like Stony Brook…
- Learn from good books…

Algorithms, Programming, Discrete Math…







For Further Reading – www.jai-tech.com