Sports Analytics in the Mainstream: What's Established, What's Emerging, and the Relationship to Legalized Sports Betting.

Presented by:

Rick Cleary
Department of Mathematics and Science
Babson College

For:

Teachers as Scholars
January 30 and February 6, 2025

OUTLINE - Day I.) Thursday 1/30/25 -Probability in the morning, statistics in the afternoon.

Introductions, background and the big picture (9:30 – 10:30)

- -Introductions of participants and a warm up problem
- -Timely topic: The sports betting industry, middle ages to present.
- -A few key ideas from probability: One page of math we need!
- -Another timely 'just for fun problem': NFL overtime.

BREAK (10:30 – 10:45 am)

The house advantage explained for several contexts (10:45 – 11:30)

- -Odds and probabilities in on-line wagers
- -Expected value calculations
- -Betting contexts: casino games, sports books, parimutuel
- -Sports betting types: Against the spread, money line, parleys and more

Case study: An expected value goes terribly wrong! (11:30 – 11:50)

- -Why do lotteries attract customers? (Loss functions and utility theory)
- -Related math ideas: Doubling strategies (The reverse lottery).

LUNCH (11:50 AM – 12:50 PM)

A quick example good for lots of ages: The Monty Hall problem with three doors … and 1000 doors (12:50 – 1)

Ranking, judging and selecting (1:00 – 1:50)

- -Live investigation: Let's have a vote on … something … that leads to a ranking.
- -Name a favorite ranking. How's it done? Is it fair?
- -Sports, colleges, cities, jobs: What isn't ranked?
- -Can bettors gain an edge by understanding how voting is done?

An Objective Ranking Attempt: The Colley Matrix Approach (1:50 – 2:40)

- -Who makes the playoffs?
- -The Colley matrix approach to rankings

Week 1 Wrap Up (2:40 – 3)

- -Lessons learned? Favorite take-aways?
- -Contest: Bet 100 hypothetical dollars and we'll see how we do.
- -Mid-course feedback and discussion of topics for week 2.

Introductions, background and the big picture.
We'll begin with participant introductions and ideas/questions about sports analytics.

Group exercise warm-up problem: At a school fundraiser, we set up a game in which players roll two fair six-sided dice and pay out the following amounts:
-If the two dice match, we pay the sum of the values, i.e 4-4 pays $8
-If the dice do not match, we pay the difference, i.e. 4-2 or 2-4 pays $2.
a.) What is the fair price to pay for this game so that on average contestants break even?
b.) What is a reasonable price to charge to make a little money for the school?
We will work on these together.

A trickier part c.) Once we establish the price, what is the probability we will lose money if we have 10 players? 20 players? 100 players?

The ideas in this warm-up problem … probability and expected value … will come up often in our class and key parts of the sports betting industry.

Timeline of the development of probability and sports analytics:

Probability as a math subject developed mathematically in the 1600's when the predecessors of what are now casinos began to form.

In the 1800's, understanding of probability was expanded tremendously when Gauss (and others) carefully explained the Law of Large Numbers.

Sports analytics really took off in the 1980s with the work of Bill James and others. At this time sports analytics was outside the mainstream of sports industry and the work was done by academics and hobbyists, and it was accessible to the public.

Then, teams began to get in the business. In 2003 the book Moneyball , by Michael Lewis, chronicled how the Oakland Athletics incorporated analytics to make baseball decisions. The book and the subsequent movie (2011) greatly increased public awareness (a positive for the field) but also meant that teams and leagues began all taking the field seriously. The public is less well informed as teams … and bookmakers … seek competitive advantage.

Group exercise: Since approximately 2000, what are some changes in sports strategy, training methods or structure that have been attributed to analytics?

In the last decade sports betting became popular (now legal in at least 38 states) and the landscape is changing quickly again!

Group exercise: Where do we get our information about sports, and how often do we hear about sports wagering?

.

Key ideas in probability … one page of math we need!

A good way to think about probability in terms of words: Probability is a function whose input is an *event* and whose output is a number between 0 and 1. We wrote P(event) = p, where $0 \leq p \leq 1$.

When we use our intuition to estimate the chances of something happening, we are doing probability, so let's all do one then summarize the guesses.

P(Celtics win the 2025 NBA Championship) = ____

Note that this probability estimation is very different than the calculation we did with the dice rolling problem!

Events are *sets* described by words, so we combine them with the operations of *intersection* (denoted by $\cap$) and *union* (denoted by U). Remember, $A \cap B$ is all of the objects in both A AND B while A U B is the set of all objects in A OR B or both! In general, intersections make sets smaller, while unions make them larger. Also for any event A, we define the *complement* of A, denoted A', as "not A", the probability that A doesn't happen. Naturally, P(A') = 1 – P(A).

P(A and B), or in set notation, $P(A \cap B)$ is the *joint* probability that A and B occur together. The *conditional* probability of A given B, written P(A | B), is the probability that A occurs if we know that B has also occurred. Essentially this reduces the entire sample space S to just the set B.

KEY FACTS: $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$.

-IF A and B are *independent*,then $P(A \cap B) = P(A) \cdot P(B)$; and $P(A|B) = P(A)$

-IF A and B are *mutually exclusive* $P(A \cap B) = 0$.

$P(A \cup B) = P(A) + P(B) – P(A \cap B)$

Group exercise: Let S (for sample space) be the set of the 8 sequences of heads and tails we can get from tossing a coin three times; i.e. S = {HHH, HHT, …, TTT}. Consider these events, which are subsets of the sample space:

A: The outcomes where the second toss was H.

B: The outcomes with at least two H's.

We compute P(A), P(B), P(A' ∩ B'), P(A ∩ B'), P(A U B), P(A|B) and explain their meaning. Let's invent events C and D so that A and C are independent, and B and D are mutually exclusive.

A sports problem where probability context matters: NFL Overtime

The 2024 Super Bowl LXVIII between San Francisco and Kansas City resulted in an overtime. San Francisco won the coin toss and chose to get the ball first. Coach Kyle Shanahan has been criticized for this choice. Let's see how the analytics department for the teams might have approached this problem.

Group exercise: Estimate P(San Fran wins | they receive to start overtime).
We will need to estimate some probabilities here, but with more time we could study historical trends to get more reliable values. (Suggestion: great project for class time between AP exam and end of school year.)

We look at a spreadsheet developed for this problem and experiment, then we conclude this section with a cool alternative to the coin toss that NFL teams might consider using.

The house advantage explained for several contexts.
Sports bets are typically expressed in terms of "odds", which are a way of looking at how much money will be returned for a winning bet.

Traditionally if we have said, "The odds on that event are 5 to 1", we are suggestion that the probability of the event occurring is 1/6.

Key fact: If the odds of an event are "k to m", we believe the corresponding probability of the event is (m/(k + m)). Notice that an "odds on favorite" means the k < m, so the probability of the event is believed to be more than one half. A bettor making a $1 wager at odds of k to m should receive $(k/m) in profit if the wager wins; or if we focus on the payoff, $((k/m) + 1). For wagers of other than a dollar, simply multiple by the size of the wager.

Group exercise: A bettor places $10 to win on a horse whose odds are 3 to 5. How much money will the bettor receive if the horse wins? How much is the profit?

Sports betting sites give these odds relative to wagers of $100. If an event is listed as +250, it means that the odds are 2.5 to 1 and a winning $100 wager will result in $250 profit. If the event is deemed 'odds on' the betting line will be listed as a negative number, so a bookmaker offering -150 on a game means that a $150 bet is needed to make a $100 profit.

Group exercise: For the NBA game between Milwaukee and Los Angeles Clippers on 3/4/24, the 'money line' bet on Milwaukee was -206, a bet on the Clippers was + 170. Let's convert these to odds and then to probabilities and then add them up.

Wait … shouldn't the probability that one or the other wins add up to 1? Yes it should!

BIG IDEA #1: BOOKMAKERS SELL MORE PROBABILITY THAN THERE REALLY IS!! As long as their money lines are set approximately correctly, this difference is their profit margin.

Another option on the game above was to bet a 'point spread' with the Bucks at -5.5, so if you bet on Milwaukee you need them to win by six or more; if you bet on the Clippers you win if they stay within five. BUT both of those bets were -110!

Group exercise: Let's do the odds and probability calculation for the point spread bet.

How about horse racing? If the internet is working, we'll find a race today but if not I have one ready from March 4 (th race at Yonkers Raceway, New York.) Actual results at https://racing.ustrottinhg.com/chart.aspx

Group exercise: Compute the 'total probability sold' on a horse race.

Key question: Can a bettor have an advantage? Sometimes, but it's rare and it takes a lot of work! Anyone who builds sophisticated probability models might be able to turn a profit… but that's bad news for regular citizens!

Expected value calculations are the most important tools for bettors. The expected value of a wager (or any other random variable) is the sum of each individual outcome times the probability of that outcome. We did this in our 'dice game' example.

Group exercise: We experiment with how expected values change when we change probability distributions in the basketball and horse racing examples. If we think our distribution is more realistic than the bookmaker's distribution, we may have an edge!

Betting contexts: casino games, sports books, parimutuel wagering

Every wager has two sides. An important thing for gamblers (and consumers of other products, and financial professionals) to consider is, "Who's on the other side?"

Classic casino games like roulette and slot machines are straightforward to evaluate. The casino always has an advantage in the long run! State lotteries, with very rare exceptions, also fit this model but generally have expected returns to players that are even worse than the casino games. (Note: Blackjack is a little trickier, see https://en.wikipedia.org/wiki/Bringing_Down_the_House_(book))

As explained above, sports books profit by selling excess probability. On a small scale, they are using the plot of https://en.wikipedia.org/wiki/The_Producers_(1967_film) They also offer many ways to make bets, and to combine bets. Basic bets include:

-Money line bets (as we've seen, these are bets where the favorite is "odds on" or "negative", and the underdog is "plus money".)

-Point spread bets where a difference in team's ability is reflected and both teams are typically about -110.

-"Over/under" bets on total points scored by two teams in a game.

-"Prop bets" on players, essential over/unders on an individual's statistics.

These events can be combined in "parlays" that involve several bets at once. These are much harder to win, but offer much greater reward. However:

BIG IDEA #2: Parlays and other exotic bets with larger payoffs are almost universally worse for the bettor in terms of expected value.

Horse racing in the United States uses a system called parimutel wagering where all of the money bet goes into a pool. The race track operator keeps a percentage of the pool, and winning betters share the rest. This can lead to some odd results:
https://www.offtrackbetting.com/results/102/vernon-downs.html

-

Group exercise: How can a horse pay more to show (where the bet pays off if the horse is first, second or third) than to place (which pays off only if the horse is first or second)?

Group exercise: Who is on the "other side" of the bet for each category of bet:
a.) casino games *b.) lotteries*
c.) sports book wagers *d.) parimutuel wagering*

There have been some instances where bettors have had an advantage, even in a lottery!

Group exercise/discussion: We take a few minutes to read the attached article "Mixing a Night Out with Probability …& Making a Fortune" by Kari Lock. What's the math lesson here? Do we think that the lottery personnel made this choice figuring nobody would notice the positive expected value for players? Or were they themselves unaware of it? And now how are sports books doing something similar (but not quite so risky)?

A quick look at some psychology: Gambling clearly appeals to people. Like many things, it can enhance an experience in moderation but can also lead to damaging results like addiction and the potential for ethical misconduct by players and officials.

Group exercise/discussion: Lotteries have been described as, "…a regressive tax on people who are bad at math." So why do people play?

Mathematical (or, more descriptively, economic) reasons given for the popularity of lotteries follow two main ideas:
1.) Utility theory: "The $5 I spent on that ticket is not useful to me, but a $50,000 prize would be!"
2.) A misunderstanding of probability and randomness that is pervasive.

The books by Kahneman (Thinking Fast and Slow) and Lewis (The Undoing Project) are highly recommended if you'd like to know more about this.

BIG IDEA #3: Sports books do spend time and resources analyzing who is likely to win, and in fact they hire independent companies to do this. But they spend as much or more analyzing how people bet, and what strategies will encourage larger (and lousier) wagers!

One strategy that some have suggested is the 'doubling strategy.' Go to the roulette wheel at a casino and place $5 on "black." If you win (with probability 18/38) you make a profit of $5. Stop and go home. If you lose, place $10 on "black", if you win you are now $5 ahead. Stop and go home. If you lose, place $20 on "black". And keep going…

Group exercise: Why is this approach like a "reverse lottery"? We also pick an amount a player might have access to and determine the probability that they go bust rather than win $5. And what if they played once a month … what's the chance they go bust within one year? Five years?

An important mathematical idea in the previous problems is that people generally badly underestimate the probability of streaks. (See my essay at https://ww2.amstat.org/mam/2010/essays/) This is a topic we may explore next week, but for today let's just do a quick experiment at www.random.org

Group exercise: In 100 tosses of a fair coin, what is the longest streak of heads and longest streak of tails that we see. Is this surprising?

Our last probability example is the famous Monty Hall Problem. We solve the problem in two different scenarios, and discuss how it relates to sports betting.

Monty Hall was the host of a TV game show called *Let's Make a Deal*. One regular feature gave a contestant a chance to choose a prize behind one of three doors, which we call A, B, C. In our version, one of the doors conceals a new car! The other two conceal goats. The candidate picks a door ... well, let's play!

Group exercise:
Our contestant chooses door _____. P(Car is behind door ____) = _____.

RC, playing Monty Hall, opens door _____ and there is a goat behind it.

P(Car is behind door ____ | Door ____ concealed a goat) = _____. Why?

If the contestant was offered a chance to switch doors at this point, should they? Answer in a few words. Then we replay the game with 1000 doors, one with a car and one with 999 goats. Does the player react differently this time? Explain in a sentence.

One more big idea before we move from probability to voting/ranking/judging:

BIG IDEA #4: Most of the 'experts' discussing bets in the media are actually employed by sportsbooks that want you to bet more! Anyone who tracks the 'best bets' of these people discover that they do not produce profitable results.

Ranking, judging and selecting

Let's collect preference order ballots on a topic of interest. Suppose next week's lunch is going to be catered and we need to choose one of five cuisines. The choices are: Chinese, Italian, Mexican, Soup/Sandwich and Thai.

Group exercise: Each attendee will rank the choices from favorite to least favorite using the first letter of the cuisine. For instance, if your preference order might be M-C-I-T-S. We collect the ballots and answer the sometimes difficult question, "Who wins?"

Group exercise: Let's list at least five reasonable ... and actually used ... ways to count the ballots.

Group exercise: What does this have to do with sports?

Group exercise: Name a favorite ranking, in sports or elsewhere. How's it done? Is it considered fair? What are the advantages and disadvantages of the process?

Nobody's perfect: Arrow's Theorem and implications

Putting it all together: In the 1950's Economist Kenneth Arrow studied elections systematically, defining criteria for fairness than any election should meet. Arrow's study of election systems led him to the big result in the field, for which he received a Nobel Prize in Economics.

Assumptions: Voters are each submitting an individual preference order for all candidates in the election. This preference order is strict, i.e. no ties; and transitive which means that if an individual prefers A to B and B to C, then they prefer A to C. (Election systems might produce a cycle, but individuals may not!)

Here's a non-technical description of these criteria:

1.) Universality: If there are n candidates, all n! preference orders are available to voters.

2.) Positive Association of Social and Individual Values: This was Arrow's description that he hoped people would vote sincerely.

3.) Independence of Irrelevant Alternatives: Suppose A, B, C and D are candidates. If A is the winner in a system, and one of B, C, D drops out, A should still win.

4.) Citizen Sovereignty: The votes are all that counts, there is no 'outside agency or rule' that favors a particular candidate. Also the system should produce a winner, with no ambiguity about tie-breaking rules.

Group Exercise: We ask 29 students who they prefer in an election with candidates D (Democrat), R (Republican) or T (Third party). The preference orders (with votes) are: D-R-T 5; D-T-R 4; R-D-T 3; R-T-D 8; T-D-R 8; T-R-D 2. If the candidates compete head-to-head in pairs: D vs R, D vs T, and R vs T, what happens?

Arrow's Theorem: With more than two candidates no election system can satisfy all of the above fairness criteria.

This is a disappointing but wonderfully comprehensive result! It's not just that the election systems we've studied are imperfect; it's that the criteria themselves are self-contradictory!

Arrow's results apply to rankings as well as to voting!

Big Idea #5: If somebody tells you that they have the 'best way' to proceed for a complicated ranking/selection problem like an award in a professional league or selection to a tournament, or judging in sports like skating or diving … you can be sure they are wrong!

What Arrow's Theorem DOESN'T say: That voting systems are bad! It says they are imperfect in that they can't meet a specified set of reasonable conditions. We can choose which conditions we think are most important/valuable in a particular situation.

There are now opportunities to wager on the results or ranking/selection/judging!

Group exercise: We research current odds on NBA Most Valuable Player for 2023-24, and how this award is determined.

Ranking in action.) Who makes the playoffs?

Twelve colleges in a region play Water Polo. They are in three four team leagues, and each team plays two games against their three league rivals. Each team also plays exactly one interleague game against a team from one of the other leagues. The results below are the in-league records of each team, followed by the interleague results.

StateLeague	W	L	Oldmac3	W	L	LibArts	W	L
Altered St.	4	2	Babton	6	0	Ambury	5	1
Deva St.	3	3	Smithley	4	2	Willity	4	2
Miss St.	3	3	SpringTech	2	4	Trinby	2	4
Enormous St.	2	4	CoastPoly	0	6	Bowherst	1	5

In the State League, Altered State beat Enormous St. twice. Each other pair of teams split. In the Oldmac 3, there were no splits, the teams always beat teams below them in the final standings. In the LibArts league, Ambury split with Bowherst; otherwise teams swept teams below them in the standings.

Interleague:

Altered St. beat Bowherst	Deva St. beat Trinby	Miss. St. beat CoastPoly
Smithley beat Enormous St.	Babton beat Willity	Ambury beat Springtech

Using the information above, pick four teams for a post-season tournament, and seed them 1 (best) to 4 (worst). Explain your reasoning in a paragraph using supporting evidence from the league standings and interleague games with math terms where applicable. Be ready to give a brief explanation you could give the coach and fans of the 'fifth best team' to explain why they were left out.

An objective approach to rankings: The Colley Matrix method
Astrophysicist Wes Colley invented a very clever method to rank teams. For several years, Colley's method was used as part of the official rankings for college football bowl selections, known as the BCS. See ColleyRankings.com

We use this on our water polo data from Fun Example 2.

Step 1.) We form a 12 X 12 matrix with twos on the diagonal, each row and column represents a team. We consider this multiplied by a vector (x_1, x_2, …, x_6), written vertically, to represent rankings of the six teams. We set this equal to a vertical vector of ones.
We take a moment here to describe matrix multiplication and see how this matrix describes a system of equations.

Step 2.) We adjust every element as follows:
Elements on the diagonal are 2 + number of games played by that team. In our example, all teams played seven games, so the diagonal elements will be nines. Off diagonal elements are -1*(number of games played between the two teams represented). Finally, the vector of ones on the right side is changed so that each value is 0.5*(2 + (wins – losses)) for the team represented.

Step 3.) To solve the system of equations we use an "inverse matrix" method which will be demonstrated in class (and I will email you all the Excel sheet.)

Excel instructions to solve linear systems: There is an easy way to solve systems of linear equations in Excel that works when the number of equations and the number of variables is exactly the same, so that the coefficients can be entered as a square matrix. When that occurs, we do the following:
-Highlight a column vector of the same size as the number of variables we are solving for. (This will be 12 in our example).
-While this is highlighted, enter the following as an equation:

=mmult(minverse(*range of coefficients),range of constants*)

where range of coefficients refers to the square matrix of coefficients, and range of constants refers to the column of constants on the right side of the system of equations.
-Then hit <CTRL>,<SHIFT>,<ENTER> all at once … this makes it a matrix operation. Don't just hit enter!!

Three things we will do:
1.) Compute rankings based on in-league games only.
2.) Re-compute using the inter-league games.
3.) Pick examples of games that won't make much difference, and those that might cause bigger variation in the rankings.
4. Back to gambling! How might we convert the Colley rankings into probabilities of winning?

The Colley method as a matrix approach sets the stage to discuss MIAA Power Rankings; that will be our first topic of week #2.

In the unlikely event that we have some time left, here are other topics that might be of interest and for which I have demonstrations at hand:

-The topology of playing fields: Open sets and closed sets.

-A key question in any comprehensive ranking system: Where's the variability? We can discuss this generally and for a great sports example we consider figure skating judges.

-"Going for it" on fourth down. We can do a process similar to our overtime example.

-Baseball as a Markov Chain. And Monopoly, see
https://www.researchgate.net/publication/266242761_Take_a_Walk_on_the_Boardwalk

Please take a few minutes to fill out the mid-term feedback form and let me know your preferences for next week.

Monty Hall Problem Solution Details:

MONTY HALL PROBLEM … Details.
Let's call the doors 1, 2, 3 and the event that the car is behind each door will be C1, C2, C3.

As we begin, P(C1) = P(C2) = P(C3) = 1/3 … car is randomized behind one of the doors, there are goats behind the other two.

Contestant picks a door, let's say C1. Host Monty Hall knows where the car is, and always shows a door with a goat to build suspense. We see Monty open door 2, let's call this event D2. Note that:

P(D2 | C1) = .5 … if the contestant has guessed the correct door, Monday randomly chooses between D2 and D3, since there is a goat behind each.
P(D2 | C2) = 0 … Monty never opens the door with the car behind it.
P(D2 | C3) = 1 … If the contestant chooses door 1 and the car is really behind door 3, Monty has no choice but to open Door 2.

What we want to know is P(C3 | D2) … given that Monty showed Door 2 is empty, what's the probability that the car is really behind Door 3?
P(C3 | D2) = P(C3 and D2)/P(D2)
= P(C3)*P(D2|C3)/[P(C1)*P(D2|C1) + P(C2)*P(D2|C2) + P(C3)*P(D2|C3)]
= ((1/3)*1) / [(1/3)*(1/2) + (1/3)*0 + (1/3)*1]
=(1/3)/(1/6 + 1/3) = (1/3)/(1/2) = 2/3 …

Similarly you can show that P(C1 | D2) is still 1/3 and it makes sense to switch!

If you prefer a simulation approach, you can experiment at
https://www.rossmanchance.com/applets/2021/montyhall/Monty.html

Compiled resources list:

Books

Albert, Jim, Teaching Statistics Using Baseball, published by Mathematical Association of America, 2003.

Cleary, Rick, "Surprising Streaks and Playoff Parity: Probability Problems in a Sports Context". Book Chapter in Joseph Gallian (Eds), *Mathematics and Sports* (pp. 16-27). Washington, DC: Mathematical Association of America. 2010. Also found on on-line by searching "Math Awareness Month 2010."

Gould, Ronald, Mathematics in Games, Sports and Gambling, published by CRC Press, 2010.

James, Bill, The Bill James Baseball Abstract 1987, published by Ballantine Books.

Kahneman, Daniel, Thinking Fast and Slow, Farrar, Straus and Giroux, 2011.

Lewis, Michael, Moneyball, WW Norton, 2004.

Lewis, Michael, The Undoing Project, WW Norton, 2016.

Mezrich, B, Bringing Down the House, Free Press, 2002

Winston, Wayne; Mathletics, published by Prinecton University Press, 2009.

Articles

Archer, A., Cleary, R., Lock, R. and Trono, J.; "March Mathness: An Analysis of a Nonstandard Basketball Pool", Math Horizons, February 2001.

Che, Yeon-Koo and Hendershott, Terrence (2009) "The NFL Should Auction Possession in Overtime Games," *The Economists' Voice*: Vol. 6 : Iss. 9, Article 5.

"The only way we wouldn't make money was if the game was fixed—in which case we'd make even more money."

Mixing a Night out with Probability... & Making a Fortune

Kari Lock
Williams College

When most people kick back for a beer in a bar, they probably don't see a lottery ticket and think "Hypergeometric Distribution!" However, two former probability students, curious as to how much the state was making, did just that. While in a bar one day, they took a Quick Draw lottery ticket (a popular bar game sponsored by the New York State Lottery) and decided to calculate the estimated payoff on the dollar. It wasn't until a year and a half later that their curiosity paid off... big time.

Let's take a hypothetical Friday night and pretend you are spending the evening out with your buddies at a local bar. On the edge of your table are Quick Draw tickets, displaying the numbers 1-80, with little boxes under each where you can check them off. You have the option of selecting anywhere between 1 and 10 of these numbers (or having your numbers randomly picked for you). There is a colorful TV screen across the room, and every five minutes, twenty new numbers randomly selected by the state appear on the screen. If you are feeling lucky this night out, you would pick your numbers, decide how much you want to bet, and then eagerly wait. As the new numbers appear on the screen, you would count the number of matches there are between those and your picks, and this number of matches would determine your payoff.

That is what you would do if you were a typical person out for a drink. However, maybe you just took a course in probability and remember problems from your homework on hypergeometric distributions, which result from sampling without replacement.

You start off with a set of size n, (the set of numbers 1–80), and out of this set you choose a subset of size m (the ten numbers you pick on your ticket). Out of the original set, there are r "successes" (the 20 numbers selected by the state), and you want to find the probability of some number x of these successes being contained in your subset (the probability of your picks having x matches with those on the screen). The payoff per dollar for each number of matches is given on your ticket, so once you have the probability of getting each possible number of matches, it is a simple matter to calculate what you really care about, the expected payoff.

The probability of getting x matches when choosing m numbers, $p(x)$, is found using binomial coefficients. We first need to compute the number of possible combinations for observing x matches, given a subset of size m. This is given by

$$\binom{r}{x}\binom{n-r}{m-x}.$$

The first term is the number of ways to choose the x successes for your subset from the r total successes, and for each of these possibilities, there are all the ways to choose the $m-x$ failures for your subset out of the $n-r$ total failures. Thus the product gives you the total number of ways to choose a subset of size m containing x matches. To find $p(x)$, we need to divide this quantity by the total number of possibilities for choosing a subset of size m (regardless of the number of matches). This is given by

$$\binom{n}{m}.$$

We thus observe

$$p(x)=\frac{\binom{r}{x}\binom{n-r}{m-x}}{\binom{n}{m}}.$$

From here, it is a simple matter of plugging in numbers... calculate $p(x)$ for each possible x, from which you can find the probability of each payoff, and then find the expected payoff for your dollar.

In a small town in New York State, when Quick Draw was introduced to the local bar, two former probability students did just this. They saw the ticket and made the connection between the lottery game and the hypergeometric distribution. Once the connection was made, one of them "just took out his old textbook, found the hypergeometric distribution, and plugged in the numbers." For example, choosing 4 numbers (referred to

in bar lingo as "4 Spot"), they simply had to compute the probability of getting 0, 1, 2, 3, or 4 matches. The probability of getting 1 match, $p(x)$, is worked out below:

$$p(1) = \frac{\binom{20}{1}\binom{60}{3}}{\binom{80}{4}} = \frac{\frac{20}{1} \cdot \frac{60 \cdot 59 \cdot 58}{3 \cdot 2 \cdot 1}}{\frac{80 \cdot 79 \cdot 78 \cdot 77}{4 \cdot 3 \cdot 2 \cdot 1}} = .433.$$

The remaining probabilities can be computed without much difficulty, resulting in $p(0) = .308$, $p(1) = .433$, $p(2) = .213$, $p(3) = .043$, and $p(4) = .003$. The game ticket in Figure 1 shows the payoff per dollar bet for each number of matches.

Thus the expected payoff can be computed by summing over the probability of each number of matches times the payoff for that number of matches. So we find the expected payoff per dollar for the 4 Spot game to be $.308(0) + .433(0) + .213(1) + .043(5) + .003(55) = \$0.59$. Since this is the payoff per dollar bet, any intelligent person will realize that odds are you will lose money playing 4 Spot Quick Draw.

However, about one year later one of these guys was back at the bar and read an advertisement for Quick Draw—an ad stating that every Wednesday in November, payoff on the 4 Spot game is doubled! Having already utilized his probability background and computed the average payoff to be about 60¢ for the dollar (as we did above), this guy saw the ad doubling this payoff, and immediately thought "Holy ****!!! They're giving away money!" And it was immediately clear that these students were going to earn more than just an "A" from their probability class.

Together they worked out the number of times they would have to play in order to ensure making a profit. When asked about his confidence prior to playing, one of them commented "I was very confident. The sample size was large enough so that the only way we wouldn't make money was if the game was fixed—in which case we'd make even more money." They had no fear going into the event, for they both had total faith in statistics, probability, and the law of large numbers.

When the bar opened at 10 am the first Wednesday in November, they were there and ready to go. From opening until the deal expired at midnight, for all four Wednesdays in November, these two guys feverishly played 4 Spot Quick Draw. Purchasing around 1500 tickets a day, they played the maximum amount of 20 games with each ticket, betting $5 a game. As they played more and more games, they started making a profit as predicted, and were able to use their winnings to keep purchasing more tickets. The only factors limiting the number of tickets they played were the printer—it took a certain amount of time for the machine to process and print out a ticket—and the actual process of cashing in the tickets. As for the colorful monitor displaying the results, it could have been turned off for all the guys cared. They were so confident in the outcome and so busy trying to maximize the number of tickets played that they didn't even pause to observe the results on the screen overhead. It turns out that their faith in probability was justified. Their final profit after the four days of playing ended up within $100 of what they had originally computed to be their expected payoff.

By the fourth Wednesday in November people were starting to realize what was going on, and several groups monopolized the Quick Draw games in bars around the county. One man, deciding to capitalize on the success of others without bothering to work out the math himself, ran into trouble. He spent the whole day playing 5 Spot Quick Draw (which didn't have the double the payoff special). I'm sure you can predict the result for this unfortunate man.

For the original two students, however, their knowledge of probability and their initiative to apply it proved to be extremely fruitful. After purchasing a new house and a new car, one of the guys was asked to comment on the experience. His words of wisdom after the whole event: "It shows that paying attention in math class can, in fact, be useful."

So exactly how rich did the combination of probability and curiosity make these guys? Well... you do the math. ■

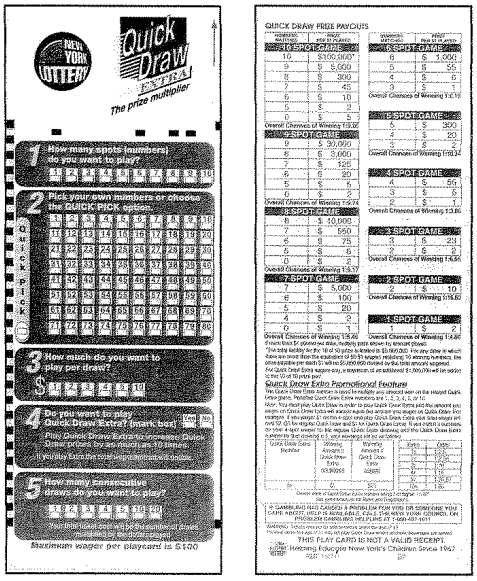

Figure 1. A New York State Lottery Quick Draw Ticket. Note the portion corresponding to the payoff amounts for the 4 Spot Game.

Mid-course feedback and Week 2 Preview … feel free to use the back!

How do you like the format and what might we do differently next week?

Please rank order the topics we could emphasize next week, from 1 (most interested) to 8 (least interested), then add a little about what you'd like to know about your top choices.

____ Sports analytics applied to roster construction and in-game strategy.

____ Calculations and examples of streaks in sports.

____ "Line moves" and how sports books react to demand.

____ Probability and psychology: How people assess and act on risk.

____ Simulation methods to forecast sports outcomes.

____ Ranking, judging and voting methods in sports and other contexts.

____ Taking this course home: Workshopping presentation ideas for your grade level.

____ Big-data results presented via statistical graphics and numerical summaries.

Do you have any other examples or areas you'd particularly like us to cover next week?

Can you suggest any activities or readings in your own experience that you think would be useful for the group?