

Class 1 *Sept. 27, 2011.*

Stats 50 / MathCS 100 - The Mathematics of Sports

Professor: T. Cover, Office: Packard 254, OH Wed, 2-4pm

TA: Leo Pekelis, Office: Seq 227, OH tba

Handouts:

- List of topics
- List of references
- Questionnaire and diagnostic exam (will not be graded)

Coreq: Stat 116

Will look at new strategies, new statistics.

And Mathematics, Statistics, and Physics of Sports:

*high jump example (high jump on moon), ball bouncing example (infinite bounces in finite time)***Sports have in common:**

- (a) competition
- (b) skill
- (c) luck

Golden age of Sports statistics and math

- (a) Cooke
- (b) Bill James
- (c) Stern
- (d) Michael Lewis

Other topics mentioned in class:

- (a) St. Petersburg Paradox - how much would you pay to play?
- (b) Stock Market and Physicists - using statistical models to take out information / make predictions.

Class 2 *Sept. 29, 2011.*Handout - Solutions to Questionnaire / Diagnostic Quiz

Note - Walt Dropo (1952), and Matt Diaz (2006) tied the record for most consecutive hits at bat in the MLB, which is 12.

Last time: Class overview, quiz

Today: Metasports, specific sports

Next time: Game Theory

Metasports (theorems about sports)

- (a) *All games are equally exciting.* For every game you can construct a function $p(t)$ which gives the probability of winning given your information up to time t .
- (b) *Dark before dawn.* It's rare for the winning team to go down to 10-1 odds of losing.
- (c) *1/2-time theorem.* If Δ_1 is the spread in the first half, and Δ_2 the spread in the second half, then given the first half spread, the chance of winning is uniformly distributed on $[0, 1]$. In other words, the game being half over doesn't preclude any chance of winning. Written in probability notation, this is $P(\Delta_1 + \Delta_2 > 0 | \Delta_1) \sim U[0, 1]$. Note, we had to assume the Δ_i 's are independent and identically distributed about 0.
- (d) *Longer games vs better players.* Had a discussion in context of golf. Longer games show erratic skill.
- (e) *λ -lengths*
- (f) *Skills vs luck*
- (g) *Simpson's paradox.* The example from class was that you can have two players, the first with a higher batting average every year for 10 years, but if you look at the overall batting average, the second will be higher.
- (h) *Game Theory.* We looked at 2x2 table payoffs from a game with two players and 1 action. We showed that sometimes making a random choice between your options is optimal. We also looked at runners and tacklers in football. What is the optimal strategy for each?

Specific Sports

- Baseball: hit velocity, bunts, curves, left hand vs right hand

Curve balls are not optical illusions! What is the radius of the circle whose arc is the trajectory of a curve ball?

Class 3 Oct. 4, 2011.

Last time: specific sports

Today: tackles, high jump, game theory

High Jump:

The high jump record on Earth was set by Javier Sotomayor from Cuba in 1993. He cleared 2.45m, or 8 ft 0 in. What would be the corresponding record on the moon? Since we expect Javier's center of gravity to

be about 3', and about 1' comes from acrobatic ability in arcing his back above his center of gravity, Javier moved his mass about 4' vertically to set the record. The moon's gravity is 1/6 that of Earth's, and hence Javier would have moved his mass a corresponding $4 * 6 = 24'$, giving a total jump of about $3 + 24 + 1 = 28'$ on the moon.

Incidentally, the women's high jump record is 6 ft 10 in, set by Stefka Kostadinova from Bulgaria in '87. And if instead of measuring absolute height, but instead the differential over how tall the athlete is, the record holder becomes Stefan Holm from Sweden, who jumped 1 ft 11 in above his height of 5-11 in 2005.

Tacklers: What is a plausible $T(R)$? The optimal direction for a tackler to move taking the position of the runner R as an input. Should we also incorporate the runner's velocity as input?

Game Theory:

We usually look at a matrix with entries a_{ij} in the i th row, j th column. The payoff player I gets from outcome ij is a_{ij} , and $-a_{ij}$ for player II. This is called a zero-sum game not because the result is 0 payout but because one player's loss is the other's gain. We had two matching guesses examples,

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad , \quad \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}$$

and found that the optimal strategies were for both players to randomize. In the first, each player would chose either option with .5 chance. In the second, the optimal ratios were 5/8 and 3/8. Note in both cases the optimal choice makes the other player indifferent between his options. This reflects a tendency for such optimal strategies to **wipe out the imperfections of the opponent**. Professor Cover mentioned losing money on playing such minimax strategies in poker.

Backward Indiction in sequential games:

- (a) Tic-Tac-Toe: pretty easy to diagram all outcomes
- (b) Chess: a finite game of perfect information, but hard/impossible to diagram all outcomes. Two approaches:
 - (a) $a(s)$ - action function gives the optimal move given the current setup of the board s .
 - (b) $\phi(s)$ - value function. A value associated with how good a current setup of the board is to you. The number, by itself has no meaning. Computers look a few moves again and optimize over possible paths.
 - (c) Be a good sport, even against computers.

General Matrix Games

- (a) First mover advantage: $\min_v \max_u f(u, v) \geq \max_u \min_v f(u, v)$
- (b) saddle points: a payoff a_{ij} that's the min of its row and max of its column.

Class 4 Oct. 6, 2011. Handout: HW Set 1 (due Thurs, Oct. 13)

Last: Highjump, Tacklers, Game Theory: Perfect Info

Today: Game Theory: Dominance, 2xn, mx2, and Fundamental theorem

Theorem: $(\max_{\min}) \max_x \min_y f(x, y) \leq \min_y \max_x f(x, y)$

proof:

$$\begin{aligned} f(x, y) &= f(x, y) \forall x \text{ and } y \\ \max_x f(x, y) &\geq f(x, y) \forall x \text{ and } y \\ \min_y \max_x f(x, y) &\geq \min_y f(x, y) \forall x \\ \min_y \max_x f(x, y) &\geq \max_x \min_y f(x, y) \end{aligned}$$

An example of the maxmin theorem is the following:

$$\begin{pmatrix} 2 & 3 & 5 \\ 4 & 7 & 1 \\ 8 & 2 & 0 \end{pmatrix}$$

The min of each row is the following vector $(2, 1, 0)$, while the max of each column is $(8, 7, 5)$. Maximizing the first vector and minimizing the second gives $\max_i \min_j = 2$ and $\min_j \max_i = 5$.

David Merrick, a theatrical producer and counterpart to Donald Trump was famous for saying "It's not enough that I should succeed - others should fail." There is a humiliation factor in losing that isn't incorporated into our characterization of games.

Dominance:

Sometimes an entire row or columns in a matrix is less than or equal to another. An optimizing player would never choose this option because another choice will always give them a better payoff. This is one way to solve for optimal strategies in zero-sum games:

Ex1.

$$\begin{pmatrix} 4 & 3 & 2 \\ 3 & 2 & 2 \\ 7 & 0 & 1 \end{pmatrix}$$

Here row 1 gives higher payoff in every column than row 2 and column 2 gives lower payoff in every row compared to column 1. Both row 2 and col 1 are dominated strategies for player 1 and 2, respectively. Crossing out these dominated strategies leave the 2x2 matrix:

$$\begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix}$$

And we can further cross out the bottom row, since the top always gives higher payoff. And so the optimal value of the games is 2 and the optimal strategy is $(r1, c3)$.

Ex2.

$$\begin{pmatrix} 2 & 0 & 1 & 4 \\ 1 & 2 & 5 & 3 \\ 4 & 1 & 3 & 2 \end{pmatrix}$$

Here col 2 dominates both columns 3 and 4, and after crossing these out, row 3 dominates row 1. The remaining 2x2 matrix doesn't have dominated strategies and the solution is to randomize so that $\max \min = \min \max$.

Ex3. (Football)

$$\begin{pmatrix} 3 & 19 \\ 6 & 2 \end{pmatrix}$$

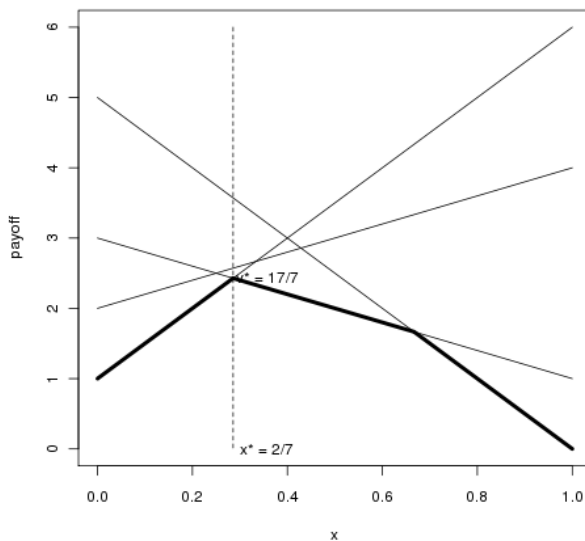
Where the payoff are the yards gained by the offense (player 1) and the choices are a pass (row 1) or run play, and for the defense (player 2) a pass (col 2) or run defense. Coaching by majority favored the offense in this game.

$2 \times n$ games:

Ex1.

$$\begin{pmatrix} 2 & 3 & 1 & 5 \\ 4 & 1 & 6 & 0 \end{pmatrix}$$

To solve these games graphically is to plot all randomizations of row 1 and row 2 as 1 line for each column, take the lower envelope (or ceiling) and find the highest point in this ceiling. The figure below shows this plot.



What happens is for any choice of randomization proportion x , meaning player 1 chooses row 1 with proportion $(1 - x)$ and row 2 with proportion x , player two can choose between the four points a vertical drawn through x intersects the randomization lines. This corresponds to choosing col 1,2,3 or 4. For any x , the optimal choice for player 2 is the lowest point. These points are exactly characterized by the lower envelope, in bold in the figure. Player 1 knowing this, will choose x^* to get the highest point on the lower envelope possible.

The solution is then for player 1 to randomize by x and for player 2 to choose a randomization of the two columns that intersect at v^* so that player 1 is indifferent between row 1 or row 2.

Notes:

- (a) Player 2 will never choose the two columns that don't intersect at v^* since they give strictly higher, or worse, payoff at x^* .
- (b) A $m \times 2$ game is solved in exactly the same way. You can either multiply all the entries by -1 or find the upper envelope.

Class 5 Oct. 11, 2011.

Today:

- Proof outline of minimax ($n \times m$)
- Duels
- Homicidal Chauffeur

Fundamental Theorem of Game Theory: Every 2-person 0-sum game has a value.

Let A be a matrix with n rows, m columns, and ij th entry a_{ij} . The possible strategies player I can choose are denoted by a vector $x = (x_1, x_2, \dots, x_n)$, $x_i \geq 0$ for all i , and $\sum_i x_i = 1$. Similarly player II chooses $y = (y_1, \dots, y_m)$, $y_j \geq 0$, $\sum_j y_j = 1$. The expected payoff, or value of the game, from playing strategies x and y is then $v = \sum_i \sum_j x_i y_j a_{ij} = x^t A y$, where the payoff is written in matrix multiplication on the right side. If you don't trust this notation, you can take any 2×2 game we've looked at and perform the multiplication to verify the right value comes out.

Theorem: $\min_y \max_x x^t A y = \max_x \min_y x^t A y = v$

Just a quick note. The inequality we proved last class is now an equality since we expand the actions each player can take by allowing mixed strategies.

proof: (Outline - all math proofs today are outlines.)

Let c_i for $i = 1, \dots, m$ denotes the columns of A , so $A = (c_1, \dots, c_m)$, and $\mathbf{C} = \{c \in \mathbb{R}^n : c = \sum_i^m y_i c_i, y_i \geq 0, \sum_i y_i = 1\}$.

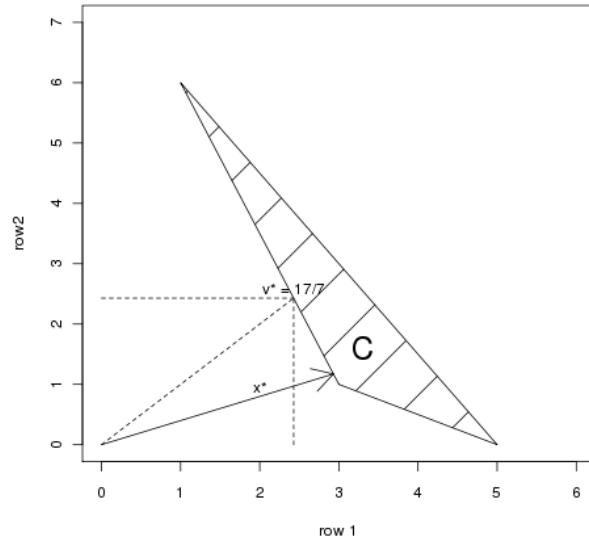
The set \mathbf{C} is called the convex hull of the vectors c_1, \dots, c_m . It is all possible linear combinations of c_i 's where the proportions add up to 1. In other words, all convex combinations of the vectors. As usual, wikipedia has a nice description of convex hulls if you're interested, http://en.wikipedia.org/wiki/Convex_hull. The figure below gives an example of \mathbf{C} in the case $n = 2$.

Since \mathbf{C} gives all of player II's strategies, player II will want to choose the strategy that is as far to the bottom-left as possible. Geometrically, this amounts to taking the top right corner of a box along the 45° line starting at the origin and finding the first point where the box touches \mathbf{C} . This point gives the value of the game and also the optimal strategy for player II. To find the optimal strategy for player I, take the unit vector going through the origin that is perpendicular to \mathbf{C} at the optimal value.

Note: In general large matrix problems are solved using linear programming, http://en.wikipedia.org/wiki/Linear_programming.

Duels

Two players start at a distance of 1 unit apart. They walk toward each other, with x denoting the distance between them. At any point, either player may choose to fire. If they hit, they win the duel. If they miss,



the players walk until $x = 0$ and the non shooting player gets to shoot with certain aim. In other words, shooting your one bullet and missing has large consequences.

Some examples of this in sports is: taking a shot at the goal in soccer (there's a large chance the keeper will take the ball out of play), and going for an uppercut in boxing.

Ex1 Suppose player 1 has accuracy function $P_1(x) = 1 - x$ and $P_2(x) = 1 - x^2$ for player 2. If you graph out these function, you can see player 2 is in general more accurate than player 1.

To solve this problem, look at it from the perspective of player 1. If player 1 decides to fire first, 1 wins with probability $P_1(x)$. If instead player 2 decides to fire first, 1 wins if 2 misses, or with probability $1 - P_2(x)$. Drawing out both curves shows a distance, past which it makes sense to fire, and before which it is better to wait. This point is found by solving $P_1(x) = 1 - P_2(x)$, giving for this problem an optimal distance of $x^* = (-1 + \sqrt{5})/2 \approx .618$.

Homicidal Chauffeur:

You're being driven around a large parking lot. You exit the car, at which point the driver decides to try to run you over. What strategy do you take to maximize your time left?

We discussed that your strategy depends on where you are in relation to the car since it has a wider turning radius than you. If you're in front of the car, run the opposite direction. If you are just inside it's turning radius, run perpendicular to the tightest arc the car can make. If you're behind the car, run win it. And if you're doing a bull herder impression with one hand on a headlight, stay to that side as close as you can.

Class 6 Oct. 13, 2011.

Handout: HW #2

Pickup: HW #1 Today:

- value of field position in football - Carter and Machol ('71), Romer ('06)

Value of field position

Let $V(x)$ be the expected score for a team at 1st and 10 when the ball is at the x yard line in football. Solving a system of equations, Romer found a curved approximation to $V(x)$. For class, we'll use $V(x) = -2 + .09x$, which is a pretty good approximation to Romer's.

Some conclusions from the equation:

- (a) every 11 yards is worth 1 point
- (b) at the 20 yard line, the expected score is 0, $V(20) \approx 0$
- (c) the cost of a turnover on 1st and 10 at the line of scrimmage is 5 points.

To see the 3rd conclusion write the cost of a turnover as $\Delta(x) = (\text{loss of your expected value}) + (\text{gain of the opponents})$ so

$$\Delta(x) = V(x) + V(100 - x) = (-2 + .09x) + (-2 + .09(100 - x)) = 5$$

A question: Are the $-2, 7$ values of $V(0)$, and $V(100)$ coincidences?

We argued yes, since having a 1000 yard field would probably wash out the influence of a touchback, and we would have $-1 * V(0) = V(1000) = 7$. Similarly on a 10 yard field, there's a big chance to score even from the 0 yard line, so we would expect $V(0) \approx 6, V(10) = 7$.

Class 7 Oct. 18, 2011.

Handout: Graded HW 1 (solution online)

Announcement: HW #2 due Tues 10/25

Last time: Field position

What if football had a 10 mi field?

- the offense would want a cross country runner
- the defense would chase with 100, 200, 400 m dashers
- it is an open problem as to how this would play out, and how should the runners pace themselves optimally.

Today:

- recurrence
- Pw vs point spread

- go for it

Go for it: you're at 4th and goal on the opponents 2 yard line

There are two options:

- (a) A field goal gives you 3 points with certainty, and assume the other team will return the kickoff to the 26 yard line. Hence the expected number of points $E(P) = 3 - V(26) = 27/11$.
- (b) Going for it will give you 7 with probability $1/3$ and a turnover $2/3$ of the time (worth $V(2)$ to the other team). Hence $E(P) = 7 * (1/3) - V(2) * (2/3) = 3.54$.

So the gain from going for it is about .9 points. Note that this is in expectation. End play generally trumps this (i.e. you're up by 6 or by 2).

Recurrence:

Ex (Coin Flips):

Let m = the expected number of flips to get heads. Then m satisfies the following recurrence relation:

$$m = p + (1 - p)(1 + m)$$

Solving for m gives $m = p^{-1}$.

Ex (Football):

Let m = the expected number of yards gained per possession. For now, assume there is only 1 down, and you either gain a yard with probability p or 0 with probability $(1 - p)$. Note 0 yards results in a turnover. Then m satisfies:

$$m = (1 - p) * 0 + p * (a + m)$$

Solving for m gives $m = a(p/q)$, where $q = 1 - p$.

Ex (Basketball):

Let m = the expected # of points per possession, $p = \text{Prob}(\text{successful shot})$ and $r = \text{Prob}(\text{Offensiverebound})$. Consider two strategies:

- (a) Always take a 2 point shot. We found m satisfies

$$m = 2p + qrm$$

and hence $m = \frac{2p}{1-qr}$. Estimating $p = .4$, $r = .2$ gives $m_2 = 10/11 \approx .91$.

- (b) Always take a 3 point shot. Following the same logic as in (a) gives $m = \frac{3p}{1-qr}$. Now estimating $p = .3$, $r = .3$ gives $m_3 = .9/.79 \approx 1.1$.

Is it possible to build a strategy around the 3 pointer with a high rebound percentage?

Probability sidenote: A random walk starting at 0 will always return to 0 in 1 or 2 dimensions, but has some chance (about 66%) of never coming back to 0 in 3 dimensions or higher. This prompted the following quote. "A drunk man will find his way home, but a drunk bird may get lost forever," - Shizuo Kakutani.

Football Pt Spread:

Define the following, p_w = probability of winning, P = point spread, m = margin of victory over point spread. And $m = F - U - P$, where F are the points scored by favorite, and U are the points scored by the underdog.

Hal Stern (paper posted in handouts section) did some data analysis to find that an approximate distribution for m is $N(.07, 13.86)$, that is m is normally distributed with mean .07 and standard deviation 13.86. We'll use slightly easier to work with number and assume $w \sim N(0, 14)$ - maybe even more accurate due to sampling variability.

From this we can calculate the probability of the favorite winning given a point spread:

$$Prob(F > U|P) = 1 - \Phi(-p/14) = \Phi(p/14) \approx .5 + .3p$$

The final approximation is due to the normal distribution function Φ being approximately linear close to 0, and is pretty accurate for $|P| < 6$.

Open question: Is more recent NFL data consistent with this approximation?

Class 8 Oct. 20, 2011.

today:

- (a) P_w vs P , pointspread in football, basketball
- (b) HORSE
- (c) QB rating

Horse

Two players take turns shooting baskets. Player I gets to start and shoot from anywhere on the court. If player I makes it, and player II does as well, the choice reverts back to player I. If, II misses, I wins (II gets a letter). If I misses, II becomes the initiator. Suppose I chooses p the probability of making a shot, and define $V = Prob(I \text{ wins})$. A recursive formula for the game is as follows where $q = 1 - p$

$$V = p^2V + pq + q(1 - V)$$

and the solution is $V = \frac{q(p+1)}{1-p^2+q}$.

- (a) Plugging in $p = .5$ gives $V = 3/5$. Can we do better?
- (b) $p = 1$ gives $V = 0 * 2/0$, so either use l'Hospital or...

- (c) let $p = 1 - \epsilon$, then $V = \frac{2\epsilon - \epsilon^2}{3\epsilon - \epsilon^2}$. As ϵ gets really small, ϵ^2 gets really really small, so $V \approx 2/3$ and $p^* = 1 - \epsilon$.

Notes:

- (a) There is a relation here to tennis where it is difficult to break an opponents serve and you have to win a set by 2 games. For instance, consider the epic 11-hour match at Wimbledon between Isner and Mahut.
- (b) You can think about another version of horse where if player I makes their shot, both players are locked into a cycle until someone misses.

Final Presentations:

- They will be in groups of 2 people.
- 20 minutes to present
- Paper in the form of handouts / outline

End quarter Calendar:

- Nov 21-25 is thanksgiving
- Nov 29: talks
- Dec 1: midterm
- Dev 6 & 8: talks

Hal Stern point spread:

Previously we saw that $M = F - U - P$ (margin over point spread) was distributed approximately $N(0, 14^2)$. For $|P| < 6$, the probability of the favorite winning is nearly linear in P , and in particular equal to $.5 + .03P$. Here's how to calculate the expected number of games a team will win given estimates of point spreads. Suppose a team will play n games, with point spreads $p = (p_1, \dots, p_n)$, and let I_i be an indicator of winning game i , i.e. 1 if game i is won, 0 otherwise. Then

$$E[N|p] = E\left[\sum_{i=1}^n I_i|p\right] = \sum_{i=1}^n E[I_i|p] = \sum_{i=1}^n \text{Prob}(F_i > U_i|p_i) = n/2 + .03\left(\sum_{i=1}^n p_i\right)$$

Draft Position: One proposal to discourage throwing games to get high draft picks is to flip the results of the last two games. In other words, LLLWLLLWW \rightarrow LLLWLLLLL. This should encourage non-playoff teams to play hard at the end of the season.

QB Rating:

Currently the QBR is $R = 100/6\left(\frac{c-30}{20} + \frac{g-3}{4} + t/5 + \frac{9.5-i}{4}\right)$. Where c is the completion %, g is the average gain per pass, t is the touchdown %, and i is the interception %. Dividing by $4a$, the number of attempts gives a clearer formula

$$R = \frac{100}{24a}(20C + G + 80T - 100I)$$

where the uppercase values are now counts instead of percentages. It is pretty clear the weights on the values are heuristically chosen and debatable on what they reflect.

There's a trade-off between quick and dirty calculations such as QBR and more sophisticated metrics. On some level ratings should fall into one of two categories, back-of-envelope calculations that can be easily calculated from box scores, and ones based on sophisticated models. There really shouldn't be anything in between.

What are some other QB ratings, ESPN's TQBR? Can we define a better QBR? What about a data driven method?

Class 9 Oct. 25, 2011.

Pickup: HW Set 2

Today:

- Composition Rule
- Bradley Terry
- Thurstone-Mosteller
- Power Index - Sagarin

Composition Rules:

Ex. A beats B with probability $2/3$, B beats C with probability $3/4$, what can we say about the probability that A beats C?

One approach is to use *Einstein's Velocity Addition Formula*. If one object is going at velocity V_1 , and a second object is going at velocity V_2 relative to V_1 , and both are going pretty fast (close to the speed of light) then the velocity of the second object from the perspective of someone standing still is no longer just the sum of the velocities. Instead it will be pretty close to $V_3 = \frac{V_1 + V_2}{1 + V_1 * V_2}$. Where we constrain the velocity to be in $[-1, 1]$.

Since probabilities are in $[0, 1]$ and we can imagine a probability of $1/2$ as being sort of like moving nowhere, one composition rule that we could try is to rescale V by $p = (V + 1)/2$ and see what the velocity addition formula gives us. This turns out to be:

$$p_3 = \frac{p_1 p_2}{p_1 p_2 + q_1 q_2}$$

where $q = 1 - p$ as usual. So does this rule make any sense, or did we just pull it out of thin air?

A second approach is to consider the *competition between city states* problem: 3 cities (A,B,C) are at war with each other. All wars are bilateral, and settled voting style, where one person is picked from the combined

population, and the home of the victor wins the war. Assuming the cities have n_A, n_B, n_C people, it is easy to see that $p_1 = \frac{n_1}{n_1+n_2}$, etc. And equivalently the odds of winning $p_1/q_1 = n_1/n_2$.

A third approach is to consider general statements like “A is twice as good as B” and “B is five times as good as C.” One way to interpret these statements is that A wins twice the number of contests that it loses when playing B, which is the same thing as the power ratio of A to B, $n_{AB} = n_A/n_B$. Since probabilities are between 0 and 1, there is only one set of probabilities (p, q) for any power ratio, and you can find it by taking $p_{AB} = n_{AB}/(n_{AB} + 1)$. And since $n_{AC} = n_A/n_C = (n_A/n_B) * (n_B/n_C) = n_{AB} * n_{BC}$, the composition rule turns out to be $p_{AC} = \frac{n_{AB}n_{BC}}{1+n_{AB}n_{BC}}$

A fourth and final approach is to think about sudden-death hockey. Assume each team scores goals as a Poisson process with rate λ_i irrespective of who they play. The distributions of any goals scored in a game between teams A and B will also be a Poisson process, now with rate $(\lambda_A + \lambda_B)$ - just add the two Poisson processes together. What’s the chance that the first goal scored will be by team A? It’s not hard to work out the math, but even intuitively, one would expect it to be the proportion of the scoring rate that team A contributes, or $\lambda_A/(\lambda_A + \lambda_B)$.

By now all these methods should start looking similar, which they should because they are all exactly the same. This is called the Bradley - Terry Model, and is often used to convert bilateral relations to indices, for example from x people preferring wine i to wine j to a wine index.

Thurston-Mosteller

Another index model to consider is Thurston-Mosteller. Assume X_i is a random variable that gives the long-run average number of points scored, normalized to have variance 1/2. By the central limit theorem, a good approximation for the distribution of X_i is $N(\mu_i, .5^2)$. Subtracting gives $W = X_i - X_j \sim N(\mu_i - \mu_j, 1)$ and the long-run chance that i wins is $P(W > 0) = \Phi(\mu_i - \mu_j)$. Backing out these probabilities is another way to solve for a composition rule: $p_3 = \Phi(\Phi^{-1}(p_1) + \Phi^{-1}(p_2))$.

Which assumptions you are prepared to accept will most likely dictate which model you like better.

Class 10 Oct. 27, 2011.

Handout: HW Set 3 (due Thurs, Nov 3rd)

Today: Finding power-indexes and the batter-pitcher problem.

A general formula for the Bradley-Terry model when there are m teams is:

$$\frac{p_{1m}}{q_{1m}} = \frac{p_{12}}{q_{12}} \frac{p_{23}}{q_{23}} \dots \frac{p_{m-1,m}}{q_{m-1,m}}$$

Solving this for p_{1m} gives (as you would expect):

$$p_{1m} = \frac{p_{12} \dots p_{m-1,m}}{p_{12} \dots p_{m-1,m} + q_{12} \dots q_{m-1,m}}$$

Note: Ralph Allan Bradley and Milton E. Terry first published their model in the paper “Rank analysis of incomplete block designs, I. the method of paired comparisons” in 1952 in *Biometrika*. Their motivating example? Taste testing pork roasts from hogs fed with corn, corn and peanuts, and corn with lots of peanuts. The results were inconclusive.

College Rankings:

One way to get college ranking of teams is to record the number of times team i beat team j as n_{ij} for every team in a league, and compute the maximum likelihood estimate (MLE) of the ranking indices $\lambda_1, \dots, \lambda_m$. In other words, assuming the Bradley-Terry model is true, meaning the outcome of a matches between teams are iid Bernoulli with team i beating team j with probability $\frac{\lambda_i}{\lambda_i + \lambda_j}$, then the probability of a matrix of outcomes (A) between teams is

$$Pr(A|\lambda_1, \dots, \lambda_m) \propto \prod_{i < j} \left(\frac{\lambda_i}{\lambda_i + \lambda_j} \right)^{n_{ij}}$$

There exist methods to find the vector of λ s that makes this probability as large as possible, which is the MLE estimate. Ford (1957) (<http://www.jstor.org/stable/2308513>) gives an algorithm to solve this, though a quick google search for “MLE for Bradley-Terry” finds papers from 2004 describing algorithms as well. I would sooner look at the later papers if I wanted to code up one of these algorithms.

Batter-Pitcher

Suppose a super batter has a BA against the league of $B = .6$, while a super pitcher has a “pitching” BA of $B_p = .1$. Finally suppose the league BA is $B_L = .25$. We can solve the question of what do we expect the BA of a super batter - super pitcher matchup to be by using the composition rule. Consider the relation

$$SB \rightarrow LP \rightarrow LB \rightarrow SP.$$

Making use of the information above we have the probability of the outcome of any adjacent matchup, i.e. $Pr(SB > LP) = .6$ and hence we can use the BT composition rule to find

$$\frac{B^*}{1 - B^*} = \frac{B}{1 - B} \frac{1 - B_L}{B_L} \frac{B_P}{1 - B_P}$$

where $B^* = P(SB > SP)$ is the batting average for the super batter compared to super pitcher, and hence $P(SB > SP) = .333$. Note that the league power ratio is inverted. This is because in the diagram we are looking at matchups of league pitchers to league batters which have a probability of not getting an on base (or winning) of $1 - B_L$.

Class 11 Nov. 1, 2011.

Handout: Solution to HW 2

Today:

- (a) Combining expert opinion
- (b) Do Longer Games favor the better player?

Combining expert opinion:

Example 1: Consider predicting whether it will rain tomorrow. There are two possible events, rain, call it event A , and no rain, call it event A^c . We have some prior information (say from the almanac) that on the average tomorrow, $P(A) = .7$. Two separate experts (weathermen) tell us that they think the probability of rain tomorrow is $P_1(A) = .75$ and $P_2(A) = .8$. What should our combined guess for the probability of raining be?

The information the experts give us can be described as $P_1(A) = P(A|X_1)$ (and the same for 2), where X_1 is the event described by the extra information expert 1 has. In other words, .75 is expert 1's best guess at the chance of rain with his information.

Our big assumption is that the experts' information are independent given A . This means that all the outside information the experts use to get their estimates doesn't overlap, or, in probability notation $P(X_1, X_2|A) = P(X_1|A)P(X_2|A)$.

The rest of the problem is a couple applications of Bayes Rule. First, say we wanted to calculate the chance expert 1's info would come up before a rainy day. The definition of conditional probability gives

$$P(X_1|A) = \frac{P(X_1, A)}{P(A)} = \frac{P(A|X_1)P(X_1)}{P(A)},$$

which we could calculate if we knew $P(X_1)$ the overall probability of getting expert 1's info. Also, an expression for the probability of rain given both experts' information is

$$P(A|X_1, X_2) = \frac{P(X_1, X_2|A)P(A)}{P(X_1, X_2)} = \frac{P(X_1|A)P(X_2|A)P(A)}{P(X_1, X_2)}$$

using the independence assumption. You'll notice we can plug in the expression for $P(X_1|A)$ from above, and we could calculate $P(A|X_1, X_2)$ from the information we have, except for the pesky $P(X_1)$, $P(X_2)$ and $P(X_1, X_2)$ values that we don't have. Here is where power ratios save us. By symmetry, we can find the same expressions for $P(X_1|A^c)$ and $P(A^c|X_1, X_2)$, just with the A 's replaced by A^c . This doesn't effect the probabilities with only X_i 's and so taking the ratio of $P(A|X_1, X_2)/P(A^c|X_1, X_2)$ will cancel all the $P(X_i)$'s and $P(X_1, X_2)$:

$$\frac{P(A|X_1, X_2)}{P(A^c|X_1, X_2)} = \frac{P(A|X_1)P(A|X_2)P(A)P(A^c)^2}{P(A^c|X_1)P(A^c|X_2)P(A^c)P(A)^2}.$$

canceling out a $P(A)P(A^c)$ from the top and bottom gives the result $\lambda_1\lambda_2P(A^c)/P(A)$ where λ_i are the power ratios, or odds, implied by the expert predictions. In general, if we had n experts, the same calculations would go through, but we would have to apply Bayes rule to $P(X_i|A)$ n times, giving a correction factor of $(P(A^c)/P(A))^{n-1}$ after canceling out the extra term from the second Bayes calculation, or as written in class

$$\frac{P(A|X_1, \dots, X_n)}{P(A^c|X_1, \dots, X_n)} = \left(\frac{P(A^c)}{P(A)} \right)^{n-1} \prod_{i=1}^n \lambda_i.$$

Longer Games:

Note there is a handout on this topic on the course website.

Consider a game with two players. Each round the distribution for the points each player can score is $X = 3$ with probability α and 0 otherwise for player I, and $Y = 2$ with probability 1 for player II. What is the chance player I wins in a contest of 1 round? This would be $P_1(X > Y) = P(X > Y) = \alpha$.

What about in two rounds? Similarly, this is $P_2(X > Y) = P(X_1 + X_2 > Y_1 + Y_2)$ where we assume each X_i and Y_i pair are independent and identically distributed according to X and Y . The only way X can win is if they score 6, which happens with a probability of α^2 .

An interesting thing is that since $\alpha^2 < \alpha$ (true for any number strictly between 0 and 1) we can be in a situation where player I is favored in a single round game, but not favored in a two round game. In particular, if we set $P_2 = 1 - P_1$, and solve the resulting equation for α , we are in a situation where player I's odds are completely reversed.

Just how bad can things get? A result proved in the handout on the website shows that for every n you can find P_1 's such that $P_n = (P_1)^n$, or $P_n = 1 - (1 - P_1)^n$, and therefore a ton of values in between.

Well at least things are monotonic, right? Meaning as the number of games increases one player will get better and better and the other worse and worse. Not necessarily. Consider John Daly, who scores better shots than Steady Eddy (or Eddie if you prefer) most of the time, i.e. 1 stroke better say 70% of the time, but 10 strokes worse 1/5th of the time (these are some very long holes). Games with less than 5 rounds will generally see John come out on top since we don't expect to see those -10 stroke games, but as soon as Eddy plays John 10 rounds, we'll expect to see about 2 bad rounds, more than canceling out John's lead in the others. Now, for arguments sake, suppose that John has a miraculous +100 round 5 % of the time. Now playing more than 20 holes is going to start looking pretty good for John since we begin to expect to see once of these miracle rounds. John is again the favorite. As you might imagine, this argument can continue forever: what if John scores -1000 1 % of the time? etc. This is what gives the oscillating pattern for $P_n(X > Y)$ seen in class.

So how much do the play time of sports determine what defines a better player?

Class 12 Nov. 3, 2011. Handouts:

(a) Graded HW # 2

(b) Longer Games

- (a) A central point of this handout is that games in which the probability of winning oscillates forever occur when $E(X - Y)$ does not exist, i.e. the expectation of both the positive and negative part are infinite. (note this is different from $E(X - Y) = \infty$.)

(c) HW Set # 4

Today: All games are equally exciting.

Define: $p(t) = Pr\{\text{Team A wins given everything we know up to time } t\}$

A way to formalize this is to define $I_A = 1$ if team A wins and 0 otherwise. Then $p(t) = E(I_A | X(t))$, $0 \leq t' \leq t$ where $X(t)$ is all the information we have up to time t .

A key claim we will use is that $p(t)$ is a martingale with respect to past information, $X(t')$, $t' \leq t$. Martingales are everywhere in probability theory and have a lot of applications, but for now the main property of a martingale you should know is that $E(p(t + \Delta) | X(t'))$, $0 \leq t' \leq t) = p(t)$. In words, the best guess of a future value of $p(t)$ from where we are standing, is exactly where we are.

A quick historical note. Martingales can be traced back to the doctoral thesis of Jean Ville, written in 1939. One of the chairs of Jean's defense was Borel (big wig in probability theory) who subsequently popularized it.

Example 1: Fly on a table.

Suppose you have a fly that buzzes around a table, starting from the center. You and your friend, on a pretty dull day, decide to make a bet. If the fly lands on your side at the end of some time, you win, and same for your friend. Say t fraction of the time has passed, and the fly is at position $X(t)$ relative to the center (positive is your side). What is the chance that you win when time runs out?

It so happens that the position of the fly $X(t)$ is a martingale. Meaning that since the fly doesn't know about your game, it has the same chance of moving up or down no matter where on the table it is. And the best guess you have for a future fly position is exactly where it's at right now.

If you further assume where the fly is in the future is normally distributed around where it is currently, you get that $X(t)$ is a Brownian Motion. It is not important to know how the calculations go from here, but we find that with these assumptions, $p(t) = \Phi\left(\frac{X(t)}{\sqrt{1-t}}\right)$.

What this says is that the end position is normally distributed about where the fly is now, with more uncertainty the more time left in the game. The figure below shows what the flight path might look like for 1 fly, and also for 100 similar flies. Notice the spread of the end point as you get closer to $t = 1$.

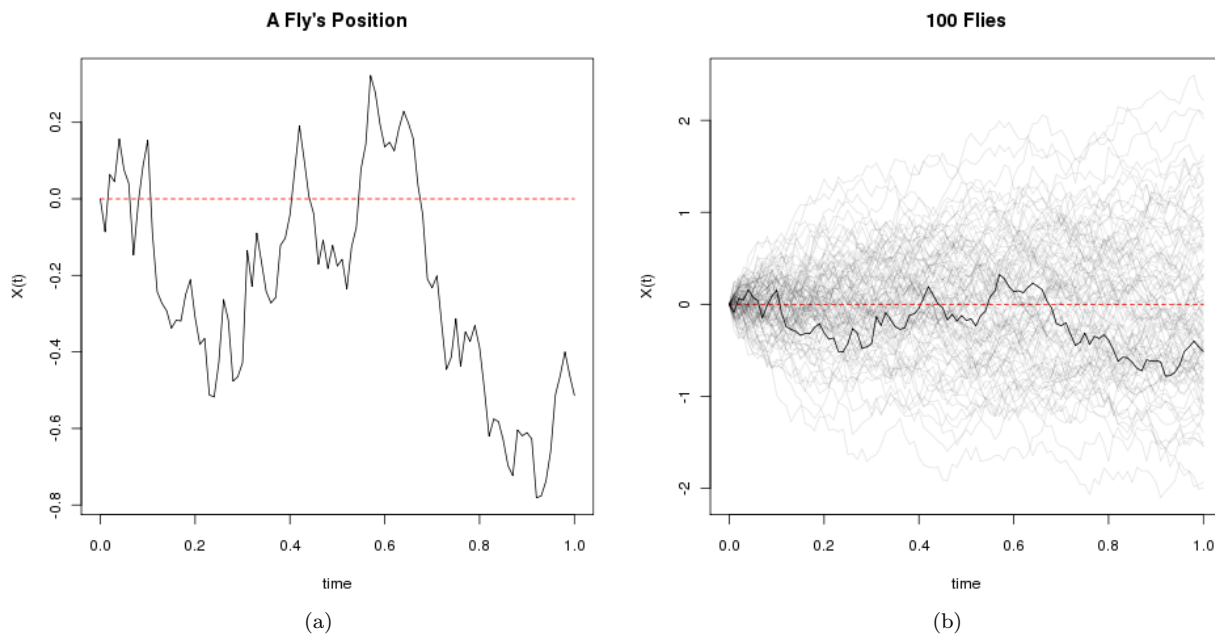


Figure 1: Flight path of 1 fly (left) and 100 flies (right).

Stopping Times: One more concept to introduce is that of a stopping time. A stopping time is a random time, T , that depends only on the past and/or present.

Example of a Stopping Time: "I will quit gambling as soon as I break even."

A few more notes about $p(t)$:

- We assume $p(t)$ is continuous, meaning the points can't jump.
- Since $E(p(1)|p(0)) = p(0)$ by the martingale property, and $p(1)$ can only be in one of two places, we can get the distribution of $p(1) = \begin{cases} 1 & : p(0) \\ 0 & : 1 - p(0) \end{cases}$.
- Say $p(0) = .5$, and we want the chance $p(t)$ hits $2/3$ before $1/3$. Since $p(t)$ will have to average out positive and negative paths, and since the distance to $2/3$ is the same as the distance to $1/3$ we can expect even odds for this to happen. But what about when it does hit $2/3$ or $1/3$?
 - Say $p(t)$ hits $2/3$ first, now the distance to 1 is the same as the distance to $1/3$, and we would expect even odds for a reversal (i.e. team A losing its advantage) before a win.
 - You can continue this logic as long as you want, and eventually you'll realize that $Pr(\geq k \text{ reversals}) = (1/2)^k$.
 - In fact, you don't even need $p(0) = .5$, having $1/3 < p(0) < 2/3$ is enough.
- In the previous point we talked about hitting discrete levels $(0, 1/3, 2/3, 1)$. There is a whole separate theory on discrete random walks. For example, say you have a walk that can either go up or down 1 step, and you keep the walk going until it either goes up to a , or down to b , then the number of steps it takes, N_{ab} has expected value $E(N_{ab}) = ab$.

–

Browsing around, I found a couple links that might be interesting on $P(t)$.

- <http://live.advancednflstats.com/> is a website you can go to and see $P(t)$ curves live (allegedly, I'll check this weekend) for NFL games.
- <http://www.advancednflstats.com/2008/08/win-probability.html> - A comment written by the owner of the site that talks about $P(t)$ a little bit, and calculates average $P(t)$ curves for various point differentials. Before you start asking how come a team that is tied at the start of 4th quarter has a 60% chance to win, these curves are for the team **with possession of the ball**, i.e. there is a 10% premium for ball possession in this situation.
- <http://www.footballcommentary.com/earlygame.htm> This post also talks about win probability. I haven't looked at it.

Game 1/2 over at 1/2 time:

If we put a few more assumptions on $P(t)$, we get the following result. Take Δ_1 and Δ_2 to be the score differential in the 1st and 2nd halves. Assume they are independent and identically distributed, with distributions that are symmetric about 0, so that $Pr(\Delta_i \leq 0) = .5$. Then *the probability that A wins, given that we are at halftime* is

$$P(1/2) = Pr(\Delta_1 + \Delta_2 > 0 | \Delta_1) = Pr(\Delta_2 > -\Delta_1 | \Delta_1) = F(\Delta_1) = F \circ F^{-1}(u) = u$$

In the above, F is the distribution function of Δ_i , and the second to last equality means that the 1/2-time score Δ_1 corresponds to some quantile of F , i.e. if $\Delta_1 = 0$, then $u = .5$. The result means that the chance team A wins at 1/2-time can be anywhere between 0 and 1 with equal probability.

Class 13 *Nov. 8, 2011.*

Handouts:

- Graded HW # 3
- HW # 3 Solutions

Today: Darkest before dawn, Physics of baseball

Darkest before dawn

How low does the probability of winning sink before a team wins? For these calculations we assume that $P(t)$ is continuous.

Aside: For the game: guess if the next card is red or black in a deck of cards, is there a stopping rule better than guess right away?

- At first thought, maybe rules like “wait until there are more black cards flipped” will give you an edge. But they turn out to be no better than random guessing since at the start all permutations of the 52 cards are equally likely, and $Pr(\text{last} = \text{red} | \text{all info up to card } i)$ is a martingale.
- This may seem counterintuitive (at least it did to me), but consider that only in 1/2 of possible deck arrangements do you even see more black cards than red.

A result we will use is that for a continuous time martingale in 1-dimension, the chance it hits a boundary of distance a to the left before it hits one of distance b to the right is $\frac{b}{a+b}$. A figure of this is below. We can use this fact twice since $Pr(p(1) = 1, \max_{0 \leq t \leq 1} p(t) \geq x) = Pr(p(t) \text{ hits } x \text{ before } 1, \text{ and hits } 1 \text{ before } 0 \text{ after hitting } x) = Pr(p(t') = x, 0 < t' < 1) Pr(p(t) = 1 | p(t') = x) = \left(\frac{.5}{1-x}\right) \left(\frac{x}{1}\right) = \frac{.5x}{1-x}$. We can break the probability into two parts since martingales only care about the past through where they are right now. In this case, conditional on $p(t')$, $p(t)$ for $t > t'$ is independent of $p(s)$ for $s < t'$. And the two fractions come from the boundary hitting result (see figure below).

Finally, the case for player II winning, or $p(1) = 0$, is exactly symmetric giving, $Pr(\text{winner falls below } x) = \frac{.5x}{1-x} + \frac{.5x}{1-x} = \frac{x}{1-x}$.

Example: If $x = 0.1$, then $Pr(\leq x) = \frac{1}{99}$ and if $x = .5$ then $Pr(\leq x) = 1$ since $p(0) = .5$ already.

Physics of Baseball:

(Bouncing balls) Do fast balls in baseball travel farther?

Imagine a pitch coming at a batter at velocity V_p and a batter swinging with velocity V_B . Then, assuming a perfect coefficient of rest, the baseball will have an exit velocity of $V_B + V_p$ with respect to the bat. Since the bat is still moving at velocity V_B in this simplified example, we get $V_{ttl} = V_B + V_p + V_B = 2V_B + V_p$.

Note that in practice (1) the exit bat velocity will be $< V_B$, and (2) the coefficient of restitution of a baseball is < 1 .

Example 2:

Consider bouncy balls instead of baseballs and gravity as the pitcher. If we drop two stacked bouncy balls on the ground, we expect a similar result. Specifically, if the downward velocity of the balls is v , and we assume a perfect coefficient of rest, then the instantaneous upward velocity when both balls hit the ground will also be v . Of course the smaller bouncy ball on top will be going at $2v$ relative to the big bouncy ball, and hence at $3v$ relative to the observer at rest. Since Newton's equation relating velocity to height under the effect of

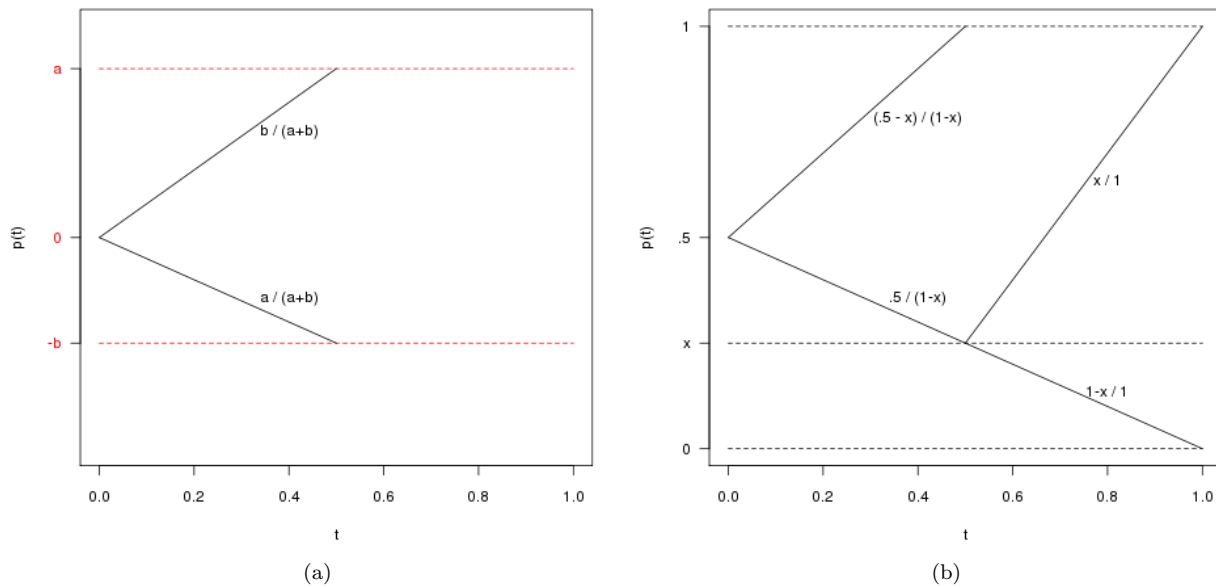


Figure 2: Prob of hitting boundary for 1-dim, continuous time martingale (left). Darkest before dawn figure (right).

gravity is $v^2 = 2gh$, the factor of 3 increase in exit velocity from from the ground will be 9 for height. This means a bouncy ball stack dropped from 5 feet should bounce the smaller ball 45 feet into the air. In class simulations supported this.

Class 14 Nov. 10, 2011.

Handouts:

- HW Set 5 (due Tues, Nov 29)
- Final Talk Signup Sheet

Today: Ball rocket, curve balls, Pythagorean Theorem, Canopy of fire, Brainstorm

Ball rocket:

Consider a stack of 3 bouncy balls. Last time we showed that the 2nd ball has an exit velocity of $3v$. Now the 3rd hits the 2nd and thus adds it's own v to the $3v$ of the 2nd, giving ball 3 an exit velocity of $4v$ relative to the second, and hence a velocity of $7v$ relative to an object at rest. Again, the effect on bounce height is proportionally squared, and hence a factor of 49 increase in height. A stack dropped from 5 ft will shoot the smallest ball 245ft into the air. No in class simulations were attempted, any individual attempts should probably be outdoors.

Canopy of fire:

Consider a cannon that can change it's shot angle, but can't move. What are all the possible points it can hit? But standard newton formula's, every cannon ball trajectory will be a parabola. A nice result is that

the upper envelope of all trajectories (the combination of all extreme points) will also be a parabola. And specifically, if the cannon can shoot a height h straight up into the air, then the parabola will have focal length h and horizontal distance $2h$.

Batted balls:

According to the exit velocity of a fastball equation we solved for last class, we should be seeing baseballs leave home plate at $V_{ttl} = 2 * 50 + 90 = 190$ mph. Why do we observe much lower speeds in real life? We have to take into account the exit velocity of the bat, now αV_B for $\alpha < 1$. And the coefficient of rest of the baseball, now $\rho(V_B + V_p)$ with $\rho < 1$. Combining these gives an updated formula,

$$V_{ttl} = (\alpha + \rho)V_B + \rho V_p$$

Example: Taking $\rho = .5$, $\alpha = .8$ gives $V_{ttl} = 110$ with a 50mph swing and a 90mph fastball. More in line with reality.

Curve balls:

What is the radius of curvature of a curve ball? The current guess of 1/2 a mile sounds too big. Another question is why do some types of curve balls seem to “break” later than others? One idea is that there are specific combinations of throw speed and angle that cause a little chaotic behavior. Another idea has to do with the ball entering a batter’s peripheral vision where the spinning distorts how the trajectory is perceived (see journal article at: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0013296>).

Bill James / Pythagorean formula:

Take p the probability of winning, or win-loss percentage, RS and RA standing for runs scored and runs allowed. The classic bill james formula is

$$p = \frac{(RS)^2}{(RS)^2 + (RA)^2}$$

Virtues?

- $0 \leq p \leq 1$
- p increases as RS increases, and decreases as RA increases

Criticisms?

- Using Taylor approximations, it seems like the chance in p should be linear in the change of RS and RA .
- What is the physical derivation?
- What is the statistical derivation?

Brainstorm:

- (a) Strategy and tactics

- (b) luck vs skill
- (c) range in baseball
- (d) extremes of sports
 - (a) bigger field (no fence in baseball)
 - (b) boxing: how many rounds is safe? should there be a weight limit?
 - (c) soccer: current pk system is unfair. What are possible solutions?
 - (d) sudden death rules / overtime
 - (e) higher basket in basketball
 - (f) ball size

Class 15 *Nov. 15, 2011.* Handouts:

- (a) OERA
- (b) Strike Zone
- (c) NERV

Today: Baseball Stats

A few questions brought up at the beginning of class:

- What is the chance Stanford wins in another matchup vs Oregon? The class answers were: 40,45,35,30,25,30,30, and 35 %.
- For rankings, do you want a loss at the beginning or end of the season?
- Would a team ever play total defense in baseball? (put all players in a ring around the batter)
- How should the mass be distributed in a shot put ball to maximize the distance it can be thrown? Currently it is homogeneous. Other options include putting all the mass around the edge (hollow), or in the center (uranium core with styrofoam cover).
- What is a good way to give golf handicaps? Currently you take the average of the lower half of your last 20 scores. The handicap is then 90 % of the difference between this and par. This makes sense between two players because he handicap forces you to play a little better than average for the top half of your game. Suppose instead there are 20 equal players in a golf tournament. Then a player would have to play in the top 5% of his game to win. By this logic, maybe handicaps should be based on the number of players in the tournament. So for a 20 person tournament, take the average of the lower 5 % of your last 20 scores (e.g. the lowest score).

Strike Zone:

The first handout (# 15), also online at [http://www.berkeley.edu/~stat50](#), shows a chart of Ted William's batting average broken down by pitch location over the strike zone.

NERV:

The second handout (# 16), constructs a Net Expected Run Value (NERV) table. The matrix shows the expected number of runs scored in each of 24 possible game situations. Averaging over 2001 data, the chart gives the average of runs scored until the end of the inning, normalized by the number of times each situation occurred during the season. If a batter comes up to the place with 0 outs and a man on first, the expectation is that .907 runs would be scored before the end of the inning.

Such a chart is useful in analyzing the value of play strategies. For example, is it ever a good idea to bunt? Lets suppose bunting always results in an out and an advancement of any men on base. Letting (x, y) denote the possible game states, with x the number of outs and y the men on base, some possible bunt transitions are $(0, 1) \rightarrow (1, 2)$, $(1, 1) \rightarrow (2, 2)$, and $(1, 2) \rightarrow (2, 3)$. The value created by the play is the NERV of the end state minus the NERV of the starting state. So the value of the three bunt transitions are $\Delta_1 = .72 - .907 = -.183$, $\Delta_2 = .347 - .544 = -.197$, and $\Delta_3 = .391 - .720 = -.329$. In each state bunting results in a negative expected run change.

OERA:

How do we evaluate baseball players? Some popular metrics are batting average, $BA = \frac{H}{AB}$, on base percentage, $OBP = \frac{H+W}{AB+W}$, and slugging average $SA = \frac{1S+2D+3T+4H}{AB}$. Although easy to calculate, these values miss out on information important for differentiating players, and contain arbitrary weightings. The OERA, or Offensive Earned-Run Average, is an attempt to present a statistic that solves these issues. It was introduced in a paper by Tom Cover and Carroll Keilers and is posted online as handout # 14.

The OERA is the number of runs a player would score in a game if he batted all positions in a lineup. The idea is that this gives a more direct evaluation of a player's worth, since at the end of the day, runs win games. It also attempts to take out team effects to give a personal rating. Note the paper also reviews a number of other baseball statistics created over the years.

There are two ways to calculate OERA. The first is called "play-by-play" calculation. Here you combine all at-bat outcomes of a player during a season. For instance a player that stikes out and then hits a single in one game, followed by an out, a homerun and another strikeout in the next game would have a combined inning of 01;040, scoring 2 runs. Stringing all of his at bats together in this manner will give a personal season of innings the batter completed personally. The OERA is then $\frac{RUNS}{INNINGS}9$.

The second way to calculate OERA is through simulation. Now imagine each batter carries around a length 6 vector $(p_0, p_B, p_1, p_2, p_3, p_4)$ where p_0 is the probability of an out, p_B probability of base on balls, and p_i the probability of a single, double, etc. If the outcome of each at bat is independent, we can simulate an inning by pulling outcomes from a hat with p vector weighings until we reach 3 outs. To get the expected number of runs, we need some more notation. Let s be the state of the game (man on 1st, 0 outs, etc.) and H the outcome of one draw from the batter hat. Then it's not hard to calculate $R(s)$, the expected number of runs scored in one time at bat, starting from a state s , given a vector of hit probabilities. What we've created is called a Markov Chain - a process that jumps from state to state, but doesn't care about more than 1 state into the past. We can use this fact to get a simple formula for $E(s)$ the expected number of runs scored beginning in state s ,

$$E(s) = \sum_{s'} p(s'|s)E(s') + R(s)$$

which simply says the expected number of runs scored in state s can be found by taking one step in the chain (the chance the batter will transition to any other state multiplied by the expected number of runs in the new state s') and keeping track of any runs scored during the step. In matrix notation, with the matrix Q containing the p transition probabilities, this is $E = QE + R$. Standard recurrence results give that the solution is $E = (I - Q)^{-1}R$, where I is a matrix with 1s on the diagonal and 0 everywhere else. The solution, E , is a vector of length 24, giving the expected number of runs in an inning starting from every possible

state of the game. And so the simulated OERA is $9E(1)$, 9 innings times the expected number of runs scored starting from $s = 1 = 0$ outs and 0 runners on base.

Class 16 *Nov. 17, 2011.* Handouts:

- (a) Graded Set 4
- (b) Soln's Set 4
- (c) Review of Topics

Announcements:

- (a) OH (Cover): 3-4 M 11/28
- (b) (no OH W 11/30)
- (c) Talks: 11/29, 12/6, 12/8 (required attendance)

Today:

- (a) Record Breaking
- (b) Scaling Laws (weight lifting, running, dimensional analysis)
- (c) Review of Topics

Record Breaking:

Examples: Jesse Owens, Babe Ruth, Bob Beamon, Sergei Bubka, Roger Bannister, Tiger Woods, Joe Dimaggio

Let X_1, X_2, \dots be independent and identically distributed with densities (so there's no chance of a tie), where X_i denotes the performance of the i th person. Let $I_i = 1$ if $X_i \geq X_j$ for all $j \leq i$ (i is the newcoming record holder), and $I_i = 0$ otherwise. Then since all the X_j have equal distribution, and hence an equal chance of being the largest, $Pr(I_i = 1) = \frac{1}{i}$. And the number of records broken, $N = \sum_{i=1}^n I_i$ has expected value $EN = \sum_{i=1}^n EI_i = \sum_{i=1}^n 1/i \approx \ln(n)$.

This would mean that records get less and less frequent as time goes on. If 1 athlete arrives every year, we would expect 2 record breaks in the first 10 years, but only about 3 more in the next 90. This isn't what we usually see in sports. One objection brought up in class was that athletes tend to improve over time. This is a valid concern, but not the direction we'll take in class today. Another issue is that the population size doesn't stay constant. For instance, suppose the population size grows at rate $e^{\alpha n}$, for $\alpha > 0$ per year. Then the probability of a record set in the i th year, since we're for now assuming all athletes have the same skills, is the chance that an athlete picked at random is from the i th year, or in math language

$$Pr(\text{record in } i^{\text{th}} \text{ year}) = \frac{e^{\alpha i}}{\sum_{j=0}^n e^{\alpha j}} = \frac{1}{\sum_{j=0}^n e^{-\alpha j}}.$$

Now, as n grows to ∞ , $\sum_{j=0}^n e^{-\alpha j} \rightarrow (1 - e^{-\alpha})^{-1}$, and so for large n (or after a lot of years), $Pr(\text{record in } i^{\text{th}} \text{ year}) \approx 1 - e^{-\alpha}$. Before you start thinking this approximation is far off, suppose for example $\alpha = .5$. By our formula,

$\sum_{j=0}^{\infty} e^{-.5j} = 1/(1 - e^{-.5}) = 2.5415$ to 4 decimal places, while $\sum_{j=0}^{14} e^{-.5j} = 2.5400$, and $\sum_{j=0}^{21} e^{-.5j} = 2.5415$. The conclusion of all this is that with a growing population, records are broken at nearly a constant rate.

Scaling Laws:

Ex. Weight Lifting

It's reasonable that the record for lifting weight should increase with the weight category of weight lifters, but how much weight lifted does an increase in lifter weight give? The formula for an object's volume is $V = cr^3$ where c is some constant, and r is the object's radius. This means that a 1 % increase in a person's radius gives about a 3 % increase in volume. Similarly the formula for an object's area is $A = c'r^2$ for some other constant c' . If we assume a person's weight is proportional to their volume, and the amount of muscle fibers determine a person's strength, which is proportional to area. Then substituting into the volume and area equations gives $Str \propto (wt)^{2/3}$. So a 10 % increase in weight should give around a 6.5 % increase in strength. It turns out that this formula fits extremely well empirically.

Ex. Running

Assuming a person's leg has length L , and the time, T , it takes for the leg to swing from backwards to forwards (make 1 stride) can be solved by considering the leg a pendulum acted on by gravity, then it is possible to solve for the effect longer legs have on a person's stride speed by solving a specific differential equation that describes the motion of a pendulum. Another, and substantially easier method is to use dimensional analysis. The idea is to write a general equation with unknown factors, and solve for the unknown by matching the dimensions (or units) of both sides of the equation. For the runner, take

$$T = m^\alpha g^\beta L^\gamma = m^\alpha \left(\frac{L}{T^2}\right)^\beta L^\gamma$$

where g is the force of gravity, m is the mass, and we substituted a relation for gravity in the second equality. The left side of the equation is in units of time. Time has no mass units, so it must be that $\alpha = 0$. Also the exponent on T on the right hand side should be 1, so $\beta = -1/2$, and this forces $\gamma = 1/2$. Plugging these values into the original formula gives $T = c\sqrt{L/g}$ where c is a constant that depends on the angle of the stride, among other things. So how fast does the runner run? Using the standard formula and substituting gives

$$V = D/T = \frac{L}{\sqrt{L/g}} = c\sqrt{gL}.$$

Hence run speed increases a square root of the amount leg length does. One disclaimer. The word "run" should really be changed to "walk" in the above since a sprinter overwhelmingly uses their own muscle to propel their leg forwards, not just gravity. But like any analysis, it's a start.

Ex. Plank's Constant

Here is one more example of the use of dimensional analysis in physics and string theory. Plank's constant is considered to define the granularity of the universe. The idea is that physical action can't take on any continuous value, but instead must be some multiple of a very small quantity. This very small quantity is defined as Plank's constant. Action relates to energy. We can use 2 other fundamental constants, namely G the gravitational constant, and c the speed of light to change the dimensions of Plank's constant in order to determine the granularity of other measurements, namely length, speed, and mass. The dimensionalities for

the three constant are:

$$\begin{aligned}G &= \frac{L^3}{mT^2} \\c &= L/T \\h &= mL^2/T\end{aligned}\tag{1}$$

where m is mass, T is time, and L is length. Then to get the smallest measurement of length, say, we set $L = c^\alpha G^\beta h^\gamma$, and normalize by setting $G = c = h = 1$, to get $L_p = \sqrt{Gh/c^3}$, and plugging in values gives $L_p = 1.6 * 10^{-33}$ cm. Similarly, $T_p = \sqrt{hG/c^5} = 5.4 * 10^{-44}$ sec and $m_p = \sqrt{hc/G} = 2.17 * 10^{-5}$ g. These numbers are thought to be the fundamental granularity of the universe. e.g. You can't find building blocks of length smaller than L_p . Of course, these results aren't perfect since m_p is about 2 micrograms. Wikipedia has some more information about these numbers, Planck Units.