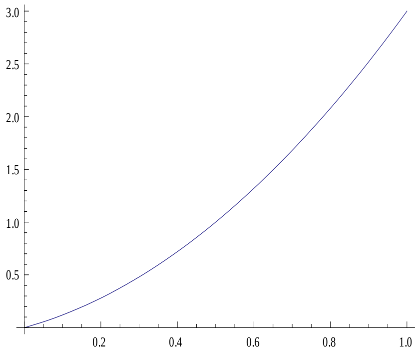
Analysing Expectation in Football

KUNAL MANDALIA

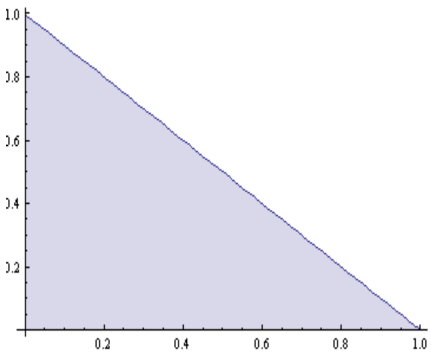
1. Motivation

Consider the function $f: [0,1] \to [0,3] : f(x) = x(2x+1)$ as shown below. The image set of f can be thought of as the point scoring system used in football. How do we interpret the domain of f? Notice $f(0) = 0, f(0.5) = 1, f(1) = 3$ so $x = 0$ indicates a loss, $x = 0.5$ indicates a draw and $x = 1$ indicates a win but what about the points in between such as $x = 0.25$?



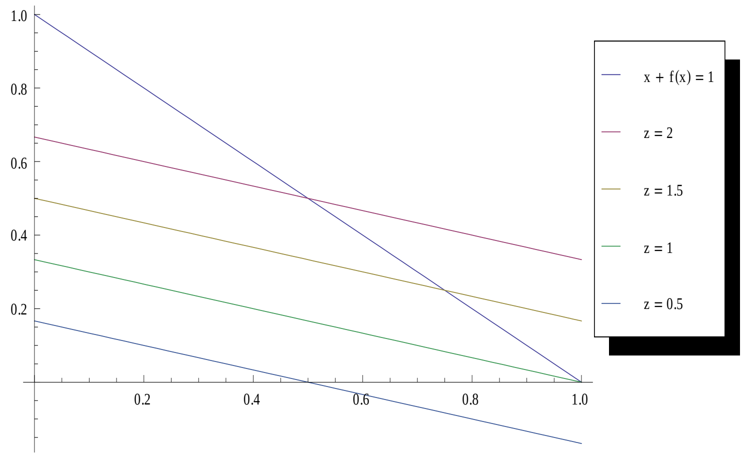
2. Deriving the expectation map

The number of points, z, a team expects to pick up against an opponent is given by the equation $z = 3(p_w) + 1(p_d) + 0(p_l) : \; p_w + p_d + p_l = 1$ thus $z = 3(p_w) + 1(p_d) : \; p_w + p_d \le 1$. Consider the function $f: p_d \to p_w : f(x) = (z - x)/3$ with $p_d, p_w \in [0,1], z \in [0,3]$. The feasible set within which the function is mapped is given by the constraint $p_w + p_d \le 1$ with $p_w, p_d \ge 0$, as shown below.



The plot below shows f given different z values.

Work in progress. Independent paper written by Kunal Mandalia
E-mail: Kunalmandalia@hotmail.com.



Notice that these functions do not all lie within the feasible set. When $z \in [1,3]$ function f intersects the line $f(x) + x = 1$ and since this line is constraining, it imposes a restriction on x, z imposes a restriction on p_d. That is to say, as the number of points a team expects to gain from a match goes from 1 to 3 they must draw fewer and fewer games, which is intuitive. When $z \in [0,1]$ the function f intersects the x axis. As z goes from 1 to 0 the team must necessarily lose more games to lower their expected number of points. Let $m: [0,3] \to [0,1]$ be the function mapping z to the maximum possible value of p_d. Given $z \in [0,1]$ we have to equate $f(x) = 0$ with $f(x) = (z - x)/3$ giving $x = z$ and given $z \in [1,3]$ we have to equate $f(x) + x = 1$ with $f(x) = (z - x)/3$ giving $x = \frac{3-z}{2}$ therefore:

$$m(z) = \begin{cases} z & : \ z \in [0,1] \\ \dfrac{3-z}{2} & : \ z \in [1,3] \end{cases}$$

Function $m(z)$ is the maximum probability of drawing a game given expected points z. Denote the set $A_z = [0, m(z)]$ the range of drawing probabilities allowed given z. The value z imposes a restriction on the range of win and loss probabilities too. Denote these B_z and C_z respectively.

The lower bound of B_z is given by solving $MIN_x \{f(x)\} = MIN_x \left\{\frac{z-x}{3}\right\} = \frac{z-m(z)}{3}$ whilst the upper bound is given by solving $MAX_x \{f(x)\} = MAX_x \left\{\frac{z-x}{3}\right\} = \frac{z}{3}$ thus $B_z = [\frac{z-m(z)}{3}, \frac{z}{3}]$.

From $p_w + p_d + p_l = 1$ we can write $p_d = x, p_w = f(x)$ thus $p_l = 1 - x - f(x)$. The lower bound of C_z is given by $MIN_x \{p_l\} = MIN_x \{1 - x - f(x)\} = MIN_x \left\{\frac{3-z-2x}{3}\right\} = \frac{3-z-2m(z)}{3}$ whilst the upper bound is given by $MAX_x \{p_l\} = MAX_x \{1 - x - f(x)\} = MAX_x \left\{\frac{3-z-2x}{3}\right\} = \frac{3-z}{3}$ therefore $C_z = [\frac{3-z-2m(z)}{3}, \frac{3-z}{3}]$. In summary:

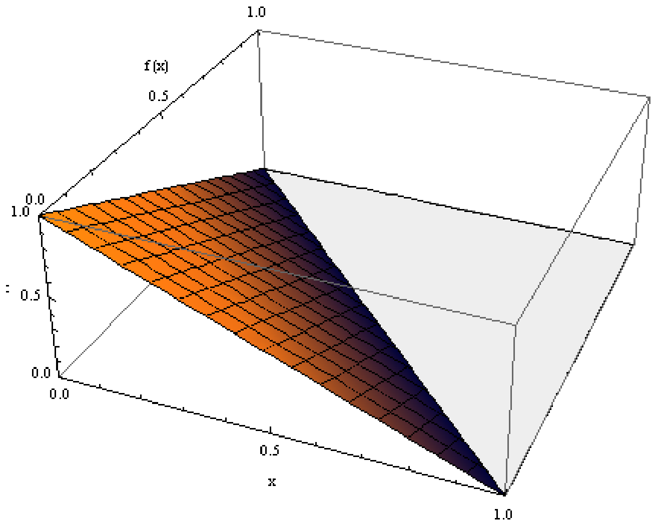
The range for a team to draw given expected number of points z is $A_z = [0, m(z)]$

The range for a team to win given expected number of points z is $B_z = \left[\frac{z-m(z)}{3}, \frac{z}{3}\right]$

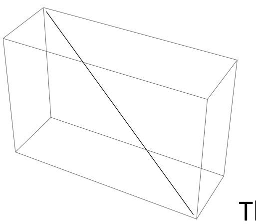
The range for a team to lose given expected number of points z is $C_z = [\frac{3-z-2m(z)}{3}, \frac{3-z}{3}]$

3. The geometry of the probability space

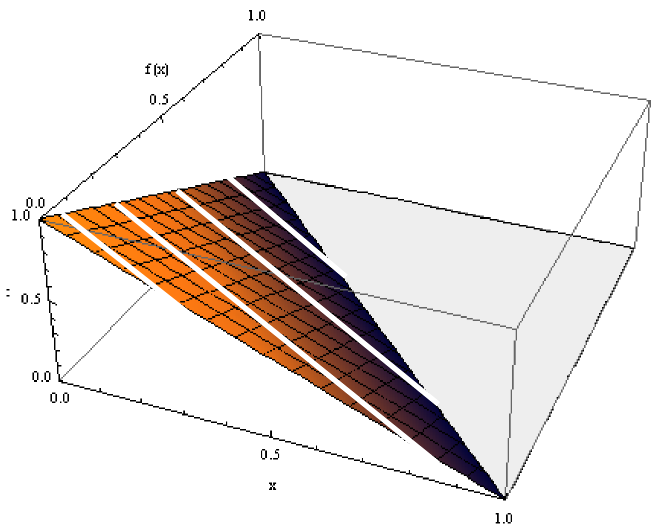
Geometrically, the space of acceptable probabilities is represented by the surface of a tetrahedron generated by equation $c + f(x) + x = 1$ where c is the probability of losing, $f(x)$ is the probability of winning and x the probability of drawing. This space is shown below.



Consider a line of constant expectation, that is, fix $z = \bar{z}$. This generates the line $f(x) = (\bar{z} - x)/3$ in the $(x, f(x))$ plane, mapping this to the surface above generates the following line in three dimensions:



This line contains all combinations of win, lose and draw probabilities that give $\bar{z}$ number of expected points. The plot below shows four such lines of constant expectation mapped onto the surface.



4. Betting on football

Given a certain number of expected points we can generate a line in three dimensions showing all probability combinations. This line does not tell us the distribution of such probability combinations, for example they may be skewed to deterministic results if a team tends to play an open game. For now let us assume that these probability combinations are uniformly distributed. Given this assumption, we can find the expected probability of winning, losing and drawing:

$$E[A_z] = E[0, m(z)] = \frac{m(z)}{2}$$

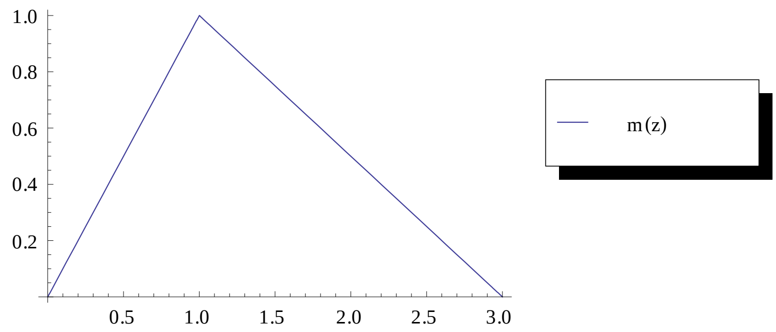
$$E[B_z] = E\left[\frac{z - m(z)}{3}, \frac{z}{3}\right] = \frac{2z - m(z)}{6}$$

$$E[C_z] = E\left[\frac{3 - z - 2m(z)}{3}, \frac{3 - z}{3}\right] = \frac{6 - 2z - 3m(z)}{6}$$

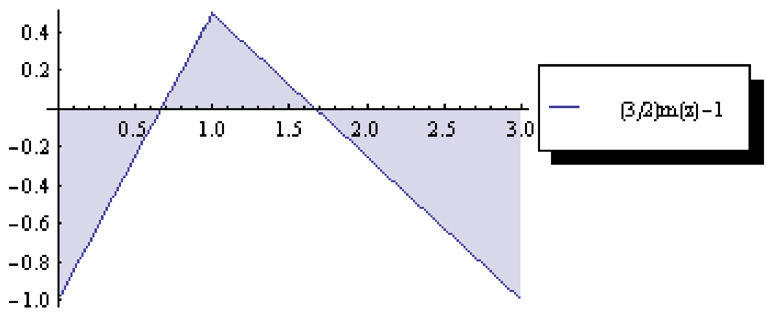
Suppose a bookie is offering 2-to-1 odds on a team to draw. If a punter takes the bet, how does his expected profit vary with z? His profit function is given by $f(p) = (2 + 1)p - 1$ for a £1 bet where p is the probability of a good result, in this case a draw.

Then $f(p) = (3)E[A_z] - 1 = \frac{3m(z)}{2} - 1$

Consider the graph of $m(z)$, shown below:

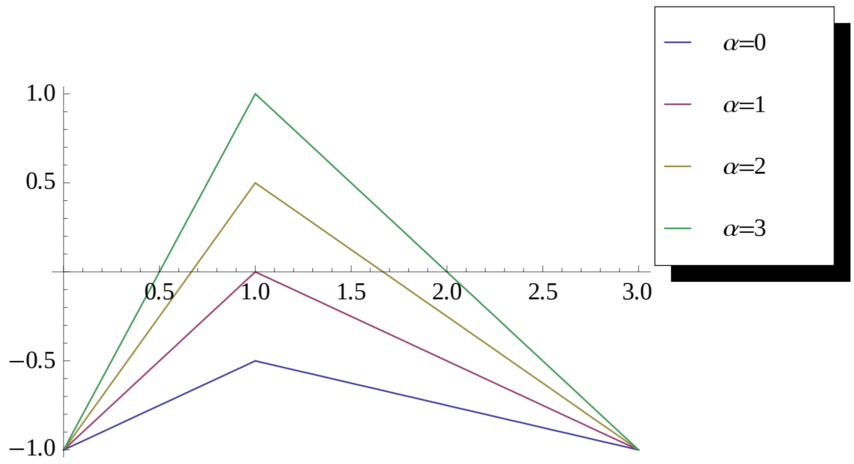


Therefore the profit function for betting on a draw will resemble a triangle:



The region above the x-axis is profit whilst the region below the x-axis is the loss incurred by the punter. The loss for the punter is the bookies profit and vice verse therefore the bookies profit function will be a reflection of the above in the x-axis.

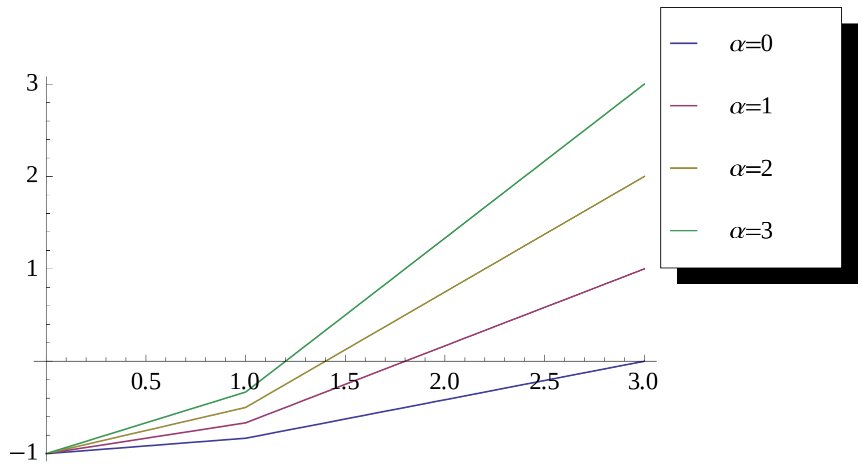
Consider the general case for when the bookie offers α-to-1 odds. The punters' profit function is given by: $f(p) = (\alpha + 1)p - 1$. The probability of a good result required to breakeven is: $p_{be} = \frac{1}{\alpha+1}$. The impact of a change in α on the profit function is highlighted below.



Suppose the punter wishes to bet on his team winning, what does his profit function look like? The profit function is given by the general equation: $f(p) = (\alpha + 1)p - 1$ with $p = E[B_z] = \frac{2z-m(z)}{6}$ thus the general profit function for betting on a win is:

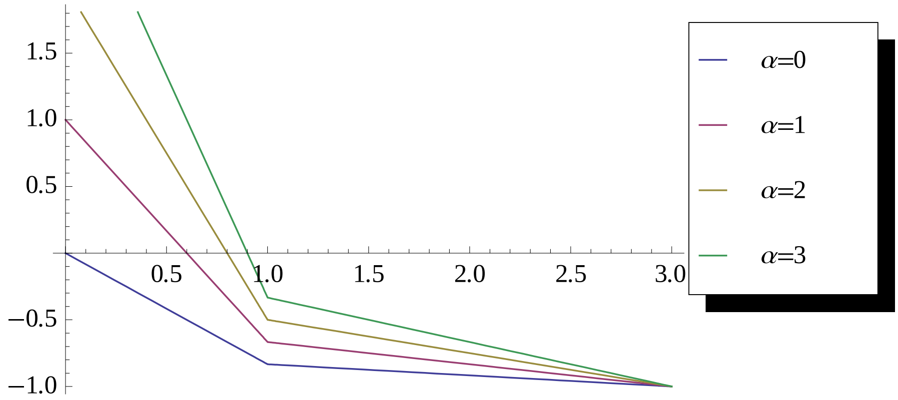
$$f(E[B_z]) = \frac{(\alpha + 1)(2z - m(z))}{6} - 1 = \begin{cases} \frac{(\alpha + 1)(z)}{6} - 1, z \in [0,1] \\ \frac{(\alpha + 1)(5z - 3)}{12} - 1, z \in [1,3] \end{cases}$$

The impact of a change in the odds offered (α) on the profit function is highlighted below.



Finally suppose the punter bets on a team to lose. The profit function for betting on a loss is given by:

$$f(E[C_z]) = \frac{(\alpha+1)(6-2z-3m(z))}{6} - 1 = \begin{cases} \frac{(\alpha+1)(6-5z)}{6} - 1, z \in [0,1] \\ \frac{(\alpha+1)(3-z)}{12} - 1, z \in [1,3] \end{cases}$$



Continuing the theme of uniform distributions across ranges A_z, B_z, C_z such that these sets are the widest possible given some z we can ask which range is the riskiest to bet on?

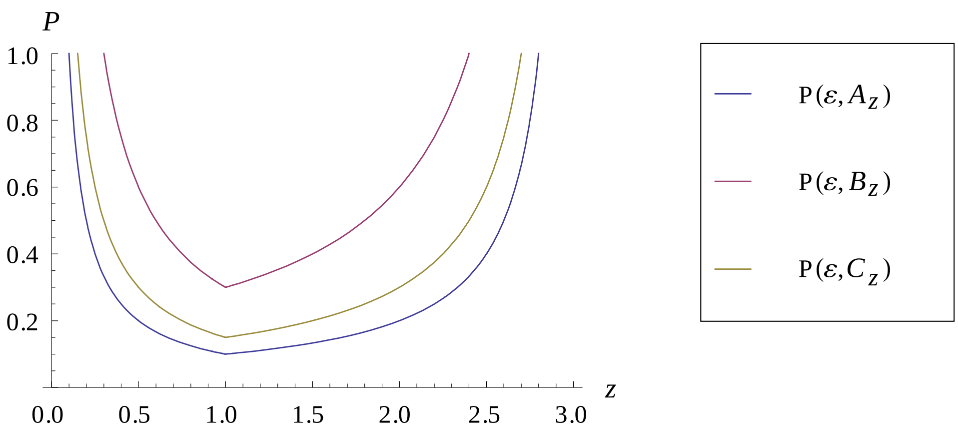
One way of assessing the riskiness of betting on a particular range is by asking what is the probability that some ε is the actual probability of an event occurring? Since we are dealing with continuous distributions let $\varepsilon = [\varepsilon_1, \varepsilon_2]$ such that $|\varepsilon|$ is small. What is the probability that the actual probability of drawing a game is in ε?

$$P(\varepsilon, A_z) = \varepsilon/|A_z| = \varepsilon/m(z)$$

$$P(\varepsilon, B_z) = \varepsilon/|B_z| = 3\varepsilon/m(z)$$

$$P(\varepsilon, C_z) = \varepsilon/|C_z| = 3\varepsilon/2m(z)$$

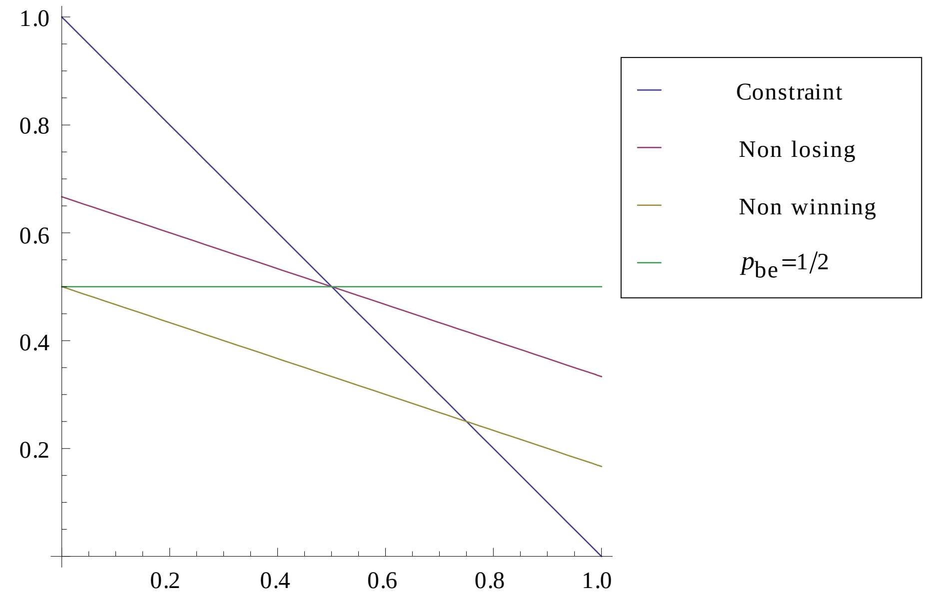
Giving the inequality $P(\varepsilon, A_z) \leq P(\varepsilon, C_z) \leq P(\varepsilon, B_z)$ where equality holds when in the limit $|\varepsilon| \to 0$. Given the assumptions above, betting on a draw represents the greatest risk and betting on a win the least, with the riskiness of betting on a loss somewhere in between, as the plot of these functions demonstrates below.



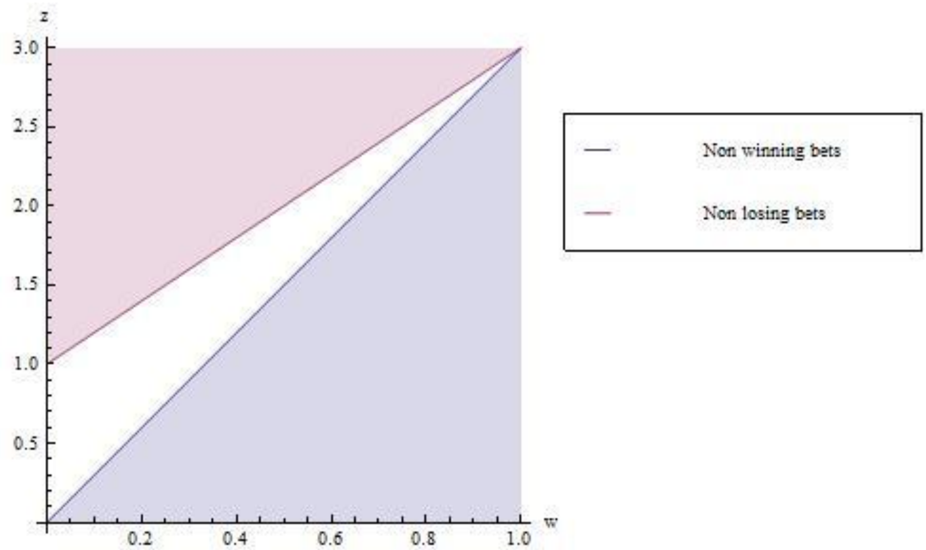
Thus far, we have been assuming uniform distribution in A_z, B_z, C_z, that is, each win, lose and draw combination is as likely as another on the three dimensional line of constant expectation. But this assumption is likely to not represent real situations very well so let us relax it and suppose that any line of constant expectation may take on any distribution, can we still infer anything using the tools developed earlier? It turns out we can.

5. Long term winning and losing bets with arbitrary distributions

Suppose a bookie is offering 1-to-1 odds on a team winning. Then the probability necessary to make this bet breakeven is $p_{be} = \frac{1}{2}$. Consider the function generated when $z = 2$, $f: [0,1] \to [0,1]$: $f(x) = \frac{2-x}{3}$. The minimum of this function occurs at (1/2,1/2) therefore $z = 2$ implies a bet would be non losing in the long run since in the worst case, the entire probability distribution is at point (1/2,1/2) which is breakeven. Therefore if $z > 2$ then the bet is profitable in the long run. Conversely consider the function generated by $z = \frac{3}{2}$, $f(x) = \frac{1.5-x}{3}$ this function intersects the breakeven line at (0,1/2) therefore $z = \frac{3}{2}$ generates a non winning probability and in the best case, the distribution of probabilities is concentrated at (0,0.5). Analogously, betting on a win when $z < \frac{3}{2}$ is a losing proposition in the long term.



Generally, let the breakeven probability of an α-to-1 bet be represented by $w = \frac{1}{\alpha+1}$. The non losing expectation associated with these odds can be derived from the intersection of function $f(x) = \frac{z-x}{3}$ and point $(1-w, w)$. Solving in terms of z gives $z = 2w + 1$. When $z > 2w + 1$ the bet is profitable for the punter in the long term. Analogously, the non winning expectation associated with these odds is found from the intersection of function $f(x) = \frac{z-x}{3}$n and point $(0, w)$ which gives $z = 3w$. When $z < 3w$ the bet is profitable for the bookie. The geometry is shown below.



[1]

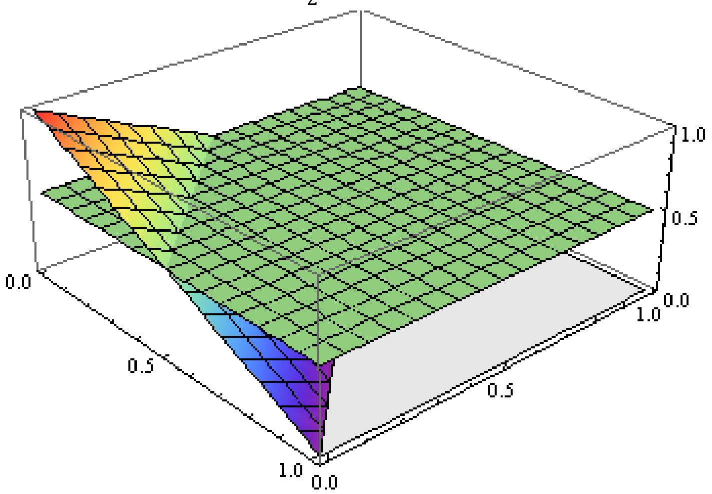
The shaded regions are where we know the long term profitability of a bet, the white region in between represents a mix between profitable and unprofitable bets which depends on the distribution of probability. It's interesting to note that despite disregarding the uniform distribution assumption, we can still say something about over 83% of the entire set of possibilities when it comes to betting on a win (The area of the white region is $\int_0^1 2x+1\,dx - \int_0^1 3x\,dx = \frac{1}{2}$ so the proportion of known area is $\frac{3-\frac{1}{2}}{3} = \frac{5}{6}$). Another point worth highlighting is what happens as the odds offered by the bookie tend to infinity.

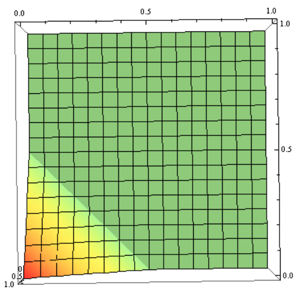
$$\lim_{\alpha\to\infty}(w) = \lim_{\alpha\to\infty}\left(\frac{1}{\alpha+1}\right) = 0$$

Non losing bets $z = 2w+1 \to 1$, this may appear unusual but the reason 1 is the minimum is because when $z \leq 1$ the line of constant expectation crosses the x-axis thus there exist the possibility of drawing all games and winning none but when $z \geq 1$ there is a positive amount of games that must be won.

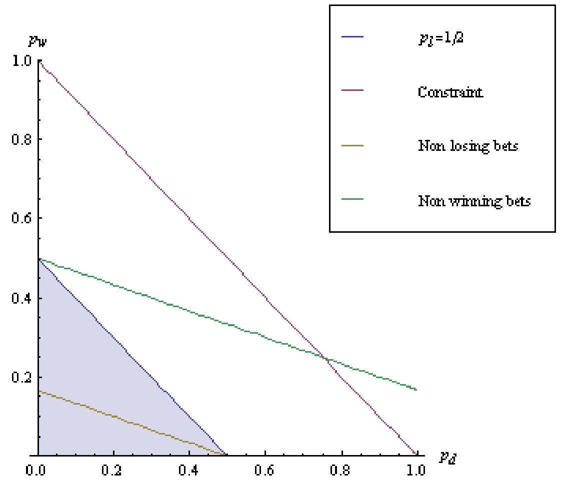
Let us now consider long term profitable bets on drawing. The probability distribution for drawing games falls in the interval $A_z = [0, m(z)]$ and since $0 \in A_z\ \forall z$ we cannot guarantee that the probability of a draw is not zero. The only certainty of drawing chances comes when we consider $z = \{0,3\}$. In either case there is always the possibility of losing, therefore, there are no long term winning or losing bets with arbitrary distributions since the profitability of a bet on a draw always depends on the distribution of A_z. In this case, by relaxing the uniform distribution assumption we cannot comment on the entire set of possibilities when it comes to betting on a draw.

Finally, consider the case of long term profitable bets on a team losing with arbitrary C_z distribution. Suppose a bookie is offering 1-to-1 odds on a team losing. By taking this bet we need the probability of losing to equal ½ in order to breakeven. Thus we have $p_l = \frac{1}{2}$ and $p_l = 1 - p_w - p_d$. The intersection of $p_w + p_d = \frac{1}{2}$ and $p_l + p_w + p_d = 1$ is shown below.

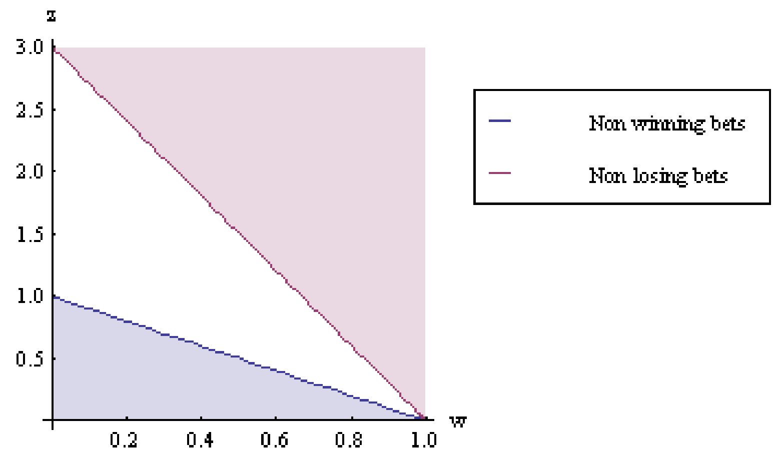




Thus $p_w = \frac{1}{2} - p_d$ and $p_d = \frac{1}{2} - p_w$ in the win-draw plane, as shown above. This line is parallel to the constraint, $p_d + p_w = 1$. The cross section in the win-draw plane is shown below. Any win-draw probability combination such that $p_d + p_w < \frac{1}{2}$ implies $p_l > \frac{1}{2}$. Consider the function $f(x) = \frac{z-x}{3}$ passing through $(0, \frac{1}{2})$ giving $z = \frac{1}{2}$. This line corresponds to non losing bets on a team losing since the smallest probability of a team losing is $p_l = \frac{1}{2}$ which is a breakeven proposition. If $z < \frac{1}{2}$ betting on a loss is profitable. Now consider the function $f(x) = \frac{z-x}{3}$ passing through $(0,\frac{1}{2})$ giving $z = \frac{3}{2}$. In the best case $p_d = 0,\ p_w = \frac{1}{2}$ thus $p_l = \frac{1}{2}$ therefore this line corresponds to non winning bets. If $z > \frac{3}{2}$ betting on a loss is unprofitable. This situation is shown below.



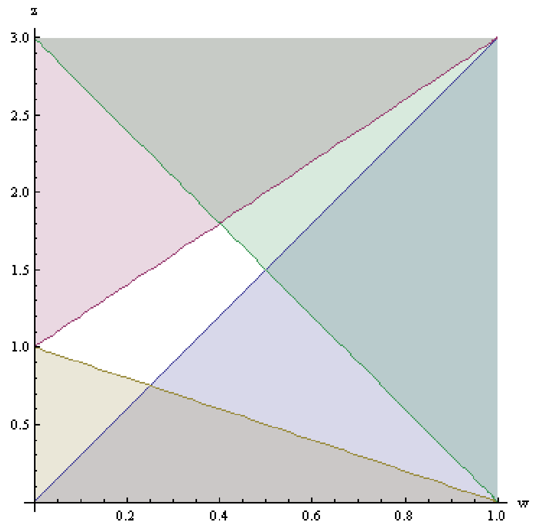
Generally, let the breakeven probability of an α-to-1 bet be represented by $w = \frac{1}{\alpha+1}$. The intersection of the surface $p_l = 1 - p_w - p_d$ with $p_l = w$ gives $p_w + p_d = (1-w)$, a line in the win-draw plane intersecting p_w and p_d at (0,1-w) and (1-w,0) respectively. The non losing function is given passing $f(x) = \frac{z-x}{3}$ through (1-w,0) giving $z = 1 - w$. The non winning function is given by passing f through (0,1-w) giving $z = 3(1-w)$. $z < 1 - w$ ensures long term profitable bets on a team losing whilst $z > 3(1-w)$ ensures long term unprofitable bets. The geometry is shown below.



[2]

We are ignorant of the long term profitability in the white region, or $\frac{1}{3}$ of the space of possibilities whilst we can comment on the remaining $\frac{2}{3}$ space of possibilities.

Consider now the intersection of plots [1] and [2], shown below.



The white quadrilateral (0,1),(1/4,3/4),(1/2,3/2),(2/5,9/5) is the region of ignorance whereas the remaining space we can say something about the LRPB.

6. The Joint expectation space

Let $P_i(p_w, p_d, p_l)$ be the probability profile for team i. Consider $P_1(x_1, x_2, x_3)$ and $P_2(y_1, y_2, y_3)$ then we have $x_1 = y_3, x_2 = y_2, x_3 = y_1$ by symmetry since the probability of team one losing is equal to the probability of team two losing, the probability of drawing is the same for both teams and finally the probability of team one losing is equal to the probability of team two winning. The number of points each team expects to pick up is given up the following:

$$E_1 = 3x_1 + x_2 \ , E_2 = 3y_1 + y_2$$

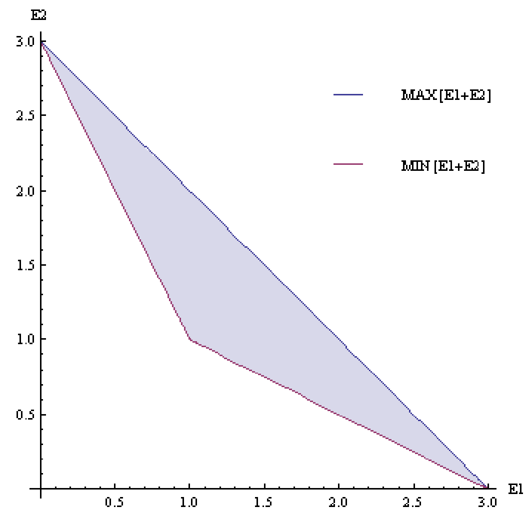
The joint or sum of both expectations is:

$E_1 +, E_2 = 3x_1 + x_2 + \ 3y_1 + y_2$. Substituting $x_2 = y_2, x_3 = y_1$ to get the joint expectation in terms of x_i gives $E_1 +, E_2 = 3x_1 + x_2 + 3x_3 + x_2 = 2(x_1 + x_2 + x_3) + x_1 + x_2$ but $(x_1 + x_2 + x_3) = 1$ thus $E_1 +, E_2 = 2 + x_1 + x_3$ and $x_1 + x_3 = 1 - x_2$ and so finally we have:

$$E_1 + E_2 = 3 - x_2$$

Joint expectation is decreasing as the proportion of draws between the teams increases, which is intuitive. We have $\max_{x_2}\{E_1 +, E_2\} = 3 : \ x_2 = 0, \quad \min_{x_2}\{E_1 +, E_2\} = 2 : \ x_2 = 1$

The geometry of the joint expectation space is shown below.



Here the line $\max_{x_2}\{E_1+, E_2\}$ is given by $f: [0,3] \to [0,3] : f(x) = 3 - x$ and the line $\min_{x_2}\{E_1+, E_2\}$ is given by $g: [0,3] \to [0,3] : g(x) = \begin{cases} 3 - 2x\,, & x \in [0,1] \\ \frac{3-x}{2}\,, & x \in [1,3] \end{cases}$

Proof:

First we will show that given $E_1 \leq 1$ the minimum joint expectation lies on $g: [0,1] \to [1,3]: g(x) = 3 - 2x$.

Let $E_1 = c \leq 1$ then the draw, win and lose probability ranges for team one are $A_c = [0, c], B_c = \left[0, \frac{c}{3}\right], C_c = [\frac{3-2c}{3}, \frac{3-c}{3}]$ respectively. From the analysis above we know that drawing the maximum amount minimises joint expectation so the minimum joint expectation for when $c \leq 1$ is given by setting $x_2 = c$. But since $E_1 = c = 3x_1 + x_2$ we have $x_1 = 0$. Therefore the probability set for team one is $P_1(0, c, 1 - c)$ and by symmetry $P_2(1 - c, c, 0)$. We need to show that the point $(E[P_1], E[P_2])$ lies on line $f: [0,1] \to [1,3] : f(x) = 3 - 2x$.

$E[P_2] = 3y_1 + y_2 = 3x_3 + x_2 = 3(1 - c) + c = 3 - 2c$ as required.

Next we show that given $E_1 \geq 1$ the minimum joint expectation lies on $g: [1,3] \to [0,1]: g(x) = \frac{1-x}{2}$. The probability ranges are now $A_c = \left[0, \frac{3-c}{2}\right], B_c = \left[\frac{c-1}{2}, \frac{c}{3}\right], C_c = [\frac{3-c}{6}, \frac{3-c}{3}]$. To minimise joint expectation, maximise x_2 so set $x_2 = \frac{3-c}{2}$ therefore $E_1 = c = 3x_1 + \frac{3-c}{2}$ giving $x_1 = \frac{c-1}{2}$ and $x_3 = 1 - x_1 - x_2 = 0$. $E_2 = 3x_3 + x_2 = \frac{3-c}{2}$ as required.

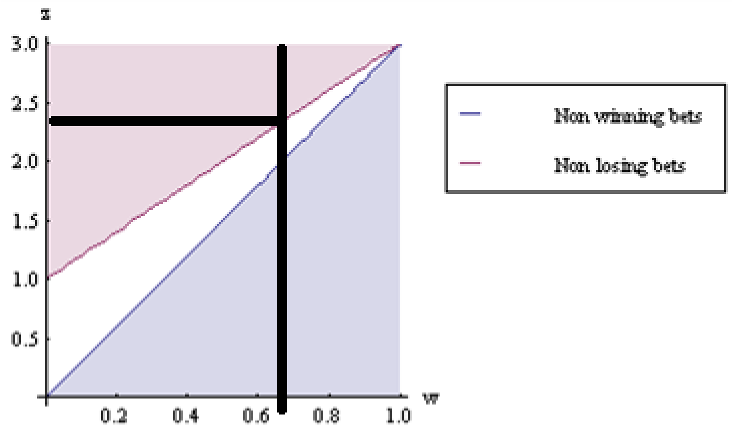
Finally we need to show that the maximum joint expectation lies on $f: [0,3] \to [0,3] : f(x) = 3 - x$. Maximum joint expectation occurs with decisive results so $x_2 = 0$. Let $E_1 = c \in [0,3]$ then $c = 3x_1$ giving $x_3 = \frac{3-c}{3}$. $E_1 = 3x_3 = 3 - c$ ■

Applied example

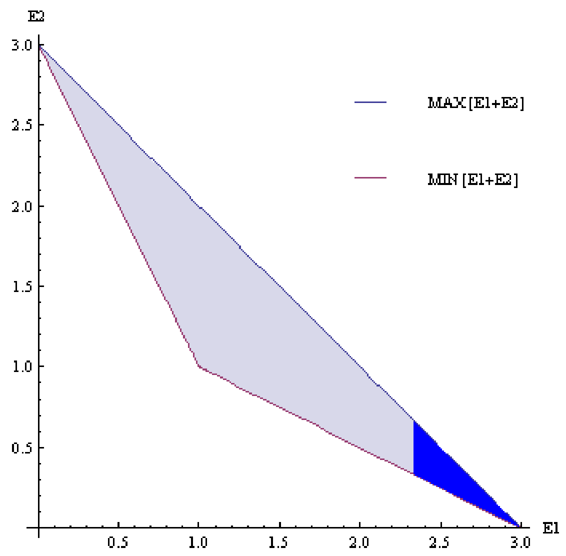
A bookie is taking bets on El Clasico: Barcelona (1.5) vs. (2.5) Real Madrid (2). That is, 0.5-to-1 for Barcelona with $\alpha = 0.5$, 1.5-to-1 for a draw with $\alpha = 1.5$ and finally 1-to-1 for Real Madrid to win with $\alpha = 1$.

Q: How many points does Barcelona need to pick up on average in order to make a bet on them a profitable one in the long term?

A: The breakeven probability is $P_{b.e} = \frac{2}{3}$. From the L.T.W.B plot we can see that $z > 2\left(\frac{2}{3}\right) + 1 = \frac{7}{3}$ would make the bet profitable.

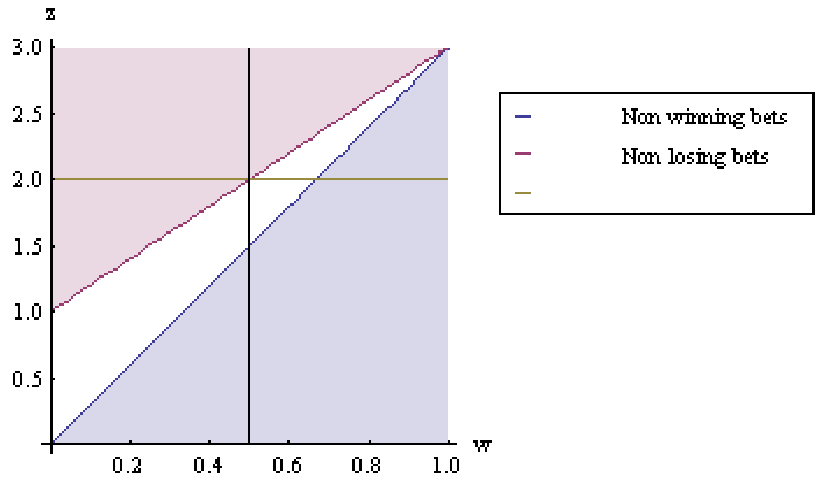


This implies the following profitable region in the joint expectation space:

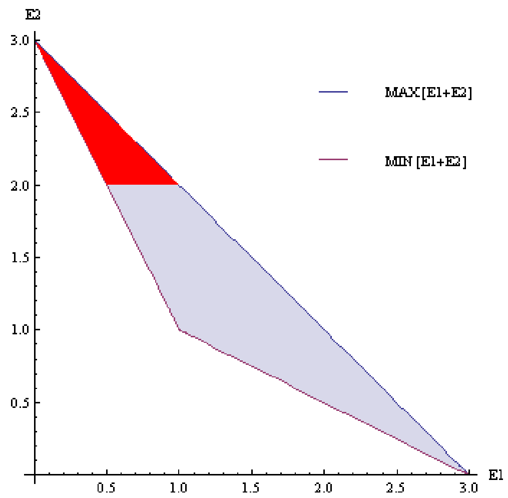


Q: How many points does Real Madrid need to pick up on average in order to make a bet on them a profitable one in the long term?

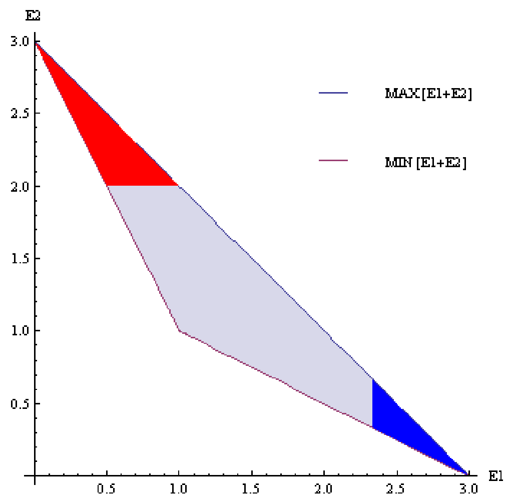
A: The breakeven probability is $P_{b.e} = \frac{1}{2}$. From the L.T.W.B plot we can see that $z > 2\left(\frac{1}{2}\right) + 1 = 2$ would make the bet profitable.



This implies the following profitable region in the joint expectation space:



Taken together, any red or blue region within the joint expectation space would represent a profitable opportunity.



7. The bookies profits

Let $X^* = (x_1^*, x_2^*, x_3^*)$ be the true probability profile for team one and the odds offered by the bookies $A = (\alpha_1 - to - 1, \alpha_2 - to - 1, \alpha_3 - to - 1)$. Let $X' = (x_1', x_2', x_3')$ represent the set of breakeven probabilities given odds A where $x_i' = \frac{1}{\alpha_i + 1}$ for $i \in \{1,2,3\}$. Suppose $x_i^* > x_i'$ for some i then there is a profitable opportunity for the punter. In competitive markets, the punters should be aware of such an opportunity and bet on it. The bookies adjust their odds to maintain a position of indifference when it comes to the outcome of an event. In particular the bookies will reduce α_i which in turn increases x_i' until the profitable opportunity is no more, that is $x_i^* \leq x_i'$, therefore this is should be a common condition in the vast majority of betting markets.

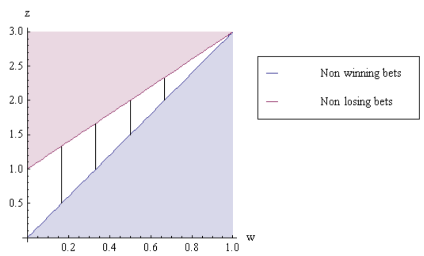
Fair markets

Consider the hypothetical market where bets on any outcome is breakeven, that is $x_i^* = x_i'$. Suppose we want to bet on a team one to win. The expected number of points team one picks up is given by $E[X^*] = 3x_1^* + x_2^*$.

Minimising expectation keeping x_1^* fixed: $\min_{x_2^*}\{E[X^*]\} = 3x_1^*$ with $x_2^* = 0$

Maximising expectation keeping x_1^* fixed: $\max_{x_2^*}\{E[X^*]\} = 2x_1^* + 1$ with $x_2^* = 1 - x_1^*$

The range of expectation for team one in a fair market is thus $R_f(E[X^*]) = [3x_1^*, 2x_1^* + 1]$. Geometrically, this range is a subset of the white region, some examples are shown below.



8. Mathematical modelling

Work in progress