The Probability Distribution of Risk Battles

By Daniel C. Taflin
November 25, 2001

Introduction

In the game of Risk, players control countries by occupying them with a variable number of "armies." The object is to gain more territory by conducting battles between neighboring countries. A battle consists of the armies of a single country going against an opponent's armies occupying a neighboring country. The battle progresses by each player rolling a prescribed number of dice, and applying rules to determine from the outcome how many armies are lost for each player. The dice are rolled repeatedly until either the defender loses all his armies (in which case the attacker can occupy the disputed country), or until the attacker is reduced to a single army (in which case there is no army to spare for occupying the opponent's country). An attacker can never lose control of his own country. The battle may also be stopped at any prior time at the attacker's discretion.

The rules for determining how many dice a player may shake are as follows:

1. The attacker may shake one less die than the number of armies on his country, to a maximum of three.
2. The defender may shake as many dice as the number of armies on his country, to a maximum of two.

The rules for deciding the outcome of a particular throw of the dice are as follows:

1. The highest attacker die is compared against the highest defender die. Whoever has the lower number loses one army. Ties go to the defender.
2. The procedure is repeated for the second-highest dice.

In cases where either the attacker or the defender only rolls a single die, a total of only one army will be lost; in all other cases a total of two armies will be lost.

Table 1. Probability Matrix for a Single Roll of Dice

	Defender dice = 1		Defender dice = 2		
Attacker dice = 1					
Attacker losses:	0	1	0	1	
Defender losses:	1	0	1	0	
Probability:	15/36	21/36	55/216	161/216	
Expected attacker losses per roll:	0.5833		0.7454		
Expected defender losses per roll:	0.4167		0.2546		
Attacker dice = 2					
Attacker losses:	0	1	0	1	2
Defender losses:	1	0	2	1	0
Probability:	125/216	91/216	295/1296	420/1296	581/1296
Expected attacker losses per roll:	0.4213		1.2207		
Expected defender losses per roll:	0.5787		0.7793		
Attacker dice = 3					
Attacker losses:	0	1	0	1	2
Defender losses:	1	0	2	1	0
Probability:	855/1296	441/1296	2890/7776	2611/7776	2275/7776
Expected attacker losses per roll:	0.3403		0.9209		
Expected defender losses per roll:	0.6597		1.0791		

Table 1 summarizes the probability of each outcome for every possible combination of attacker vs. defender dice. The following conclusions can be drawn:

1. When both attacker and defender have a large number of armies, the attacker will, on average, lose armies at a 15% slower rate than the defender.
2. Towards the end, when either the attacker or the defender must shake fewer dice, the advantage swings more strongly toward the player with the most armies.

Probability of an Attacker Victory

The course of a battle traces out a random-walk path through the two-dimensional Attacker-Defender space, ending, if taken to its conclusion, either with zero defender armies (attacker victory) or a single attacker army (attacker defeat). The probability distribution function for the path is a typical discrete diffusion problem with, however, special conditions prevailing near the boundaries where the numbers of dice thrown per roll change. In the following sections a solution is given first to the "free-space" region in which the attacker throws three dice and the defender two, then to the boundary regions, and then finally a comprehensive solution stitches these two regions together. Later sections will address the probability of an attacker successfully carrying out a sweep through a string of adjacent countries.

(a) Free-Space Solution

When the number of attacker armies, A, is greater than 3 and the number of defender armies, D, is greater than 1, the roll will be 3 attacker dice vs. 2 defender dice, and there will be three possible outcomes: attacker loses 2 armies, each loses 1 army, or defender loses 2 armies. In the attacker-defender space, this can be viewed as shown in Figure 1. After two rolls, there are five possible locations. Since a total of two armies are always lost per shake, the possible outcomes will fall on the diagonal line corresponding to points equidistant from the starting point, the distance being twice the number of rolls. Distance, here, refers to the total number of horizontal and vertical steps between points.

Viewed as a diffusion problem, diffusion actually occurs in only one dimension, in the direction from upper left to lower right. In the direction perpendicular to this, from upper right to lower left, the motion is completely deterministic, marching two steps per roll. The equation of motion for $P_{m,n}(s)$, the probability of reaching the point (m, n) in the A-D plane after s rolls of the dice, is:

Figure 1. The possible states of the system after one and two rolls.

$$(1) \qquad P_{m,n}(s) = f_{32}^0 P_{m,n+2}(s-1) + f_{32}^1 P_{m+1,n+1}(s-1) + f_{32}^2 P_{m+2,n}(s-1)$$

where f^i_{mn} is the probability of the attacker losing i armies in a roll of m dice vs. n dice.

By defining a new function $R_k(s) = P_{m,n}(s)$ where $k = m - n$ this equation can be reduced to a discrete diffusion equation in one dimension:

$$(2) \qquad R_k(s) = f_{32}^0 R_{k-2}(s-1) + f_{32}^1 R_k(s-1) + f_{32}^2 R_{k+2}(s-1)$$

However, this step turns out to be unnecessary because the solution for $P_{m,n}$ can be obtained directly by counting the number of paths between the starting point and the point of interest. In the first place, the starting point (M, N) and the endpoint (m, n) are subject to the following constraints:

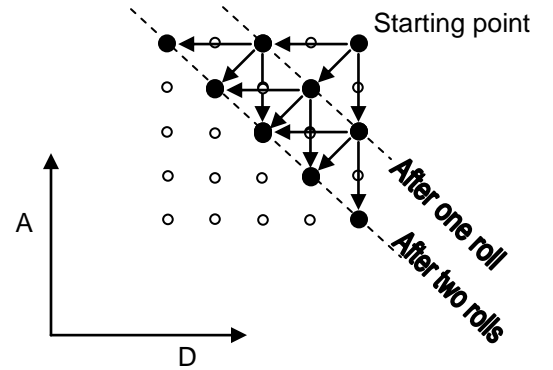- $m \le M$ and $n \le N$
- $(M + N) - (m + n) = 2s$

Now define $L = (m - n) - (M - N)$ to represent the relative advantage picked up by the attacker after s rolls. This advantage is equal to twice the excess of "good" rolls (where the attacker loses 0 and the defender loses 2) over "bad" rolls (where the attacker loses 2 and the defender loses 0). If we represent the number of "good" rolls by G, "bad" rolls by B, and neutral rolls by I (for "indifferent"), then we have the relations

(3)
$$L = 2(G - B)$$
$$s = G + B + I$$

The probability of getting from (M, N) to (m, n) is the sum of probabilities for each path between these points. These paths are each described by a triplet (G, B, I); the probability of tracing out any particular one of these paths is

(4) $$\left(f_{32}^0\right)^G \left(f_{32}^1\right)^I \left(f_{32}^2\right)^B$$

The number of paths corresponding to this triplet is the number of ways to split s items into three groups of G, B, and I members, respectively, which is simply the multinomial coefficient $(s!)/(G!\ B!\ I!)$. Thus the total probability of getting from (M, N) to (m, n) is this multinomial coefficient, multiplied by the single-path probability (4), and summed over all combinations of (G, B, I) that satisfy the constraints (3). These combinations can be enumerated as follows. Start with the minimum value of I (either zero or one depending on whether $(s - L/2)$ is even or odd), and increase I by two while decreasing G and B each by one (since the difference between G and B must remain constant) until reaching the maximum value of I, which will occur when either G or B reaches zero. For example, if $L = 4$ and $s = 12$, the allowed combinations of (G, B, I) are $(7, 5, 0)$, $(6, 4, 2)$, $(5, 3, 4)$, $(4, 2, 6)$, $(3, 1, 8)$, and $(2, 0, 10)$.

The final solution can be written:

(5) $$P_{m,n}(s) = \begin{cases} \delta_{m,M+N-n-2s} \displaystyle\sum_{0 \le k \le (s-L/2)/2} \frac{s!}{(k!)(L/2+k)!(s-L/2-k)!} \left(f_{23}^0\right)^{L/2+k} \left(f_{23}^1\right)^{s-L/2-2k} \left(f_{23}^2\right)^k, & L \ge 0 \\[2em] \delta_{m,M+N-n-2s} \displaystyle\sum_{0 \le k \le (s-L/2)/2} \frac{s!}{(k!)(L/2+k)!(s-L/2-k)!} \left(f_{23}^0\right)^k \left(f_{23}^1\right)^{s-L/2-2k} \left(f_{23}^2\right)^{L/2+k}, & L < 0 \end{cases}.$$

(b)   Boundary Region Solution

In the boundary region $(M < 4$ or $N < 2)$ the number of dice thrown depends on the number of armies. The solution for an arbitrary (M, N) in this region can be obtained by solving for $P_{m,n}$ in strips of constant M or N, beginning at the edges and working inward. Denoting the probability of an attacker victory by $Q_b(m, n)$, some simple cases can be dispensed with immediately:

(6) $$Q_b(2,1) = f_{11}^0$$

(7) $$Q_b(m,1) = 1 - \left(f_{31}^1\right)^{m-3}\left(f_{21}^1\right)\left(f_{11}^1\right), \quad m \ge 3$$

(8) $$Q_b(2,n) = \left(f_{12}^0\right)^{n-1}\left(f_{11}^0\right), \quad n \ge 1$$

Equation (7) was obtained by the observation that the only possible losing path is for the attacker to lose every single roll of the dice; the probability of winning is one minus the probability of losing. Equation (8) gives the probability for the single possible winning path. For values of m greater than 2 and n greater than 3 the solutions become more involved, but are still summable. The equation of motion for $m = 3$ and $n \ge 2$ is:

(9) $Q_b(3,n) = f_{22}^1 Q_b(2,n-1) + f_{22}^0 Q_b(3,n-2)$

Plugging in the solution for Q(2, n-1) from equation (8) gives:

(10) $Q_b(3,n) = f_{22}^0 Q_b(3,n-2) + \left(f_{22}^1\right)\left(f_{12}^0\right)^{n-2}\left(f_{11}^0\right)$

After multiplying this equation by $(f^0{}_{22})^{-n/2}$ it can be rearranged into the following form:

(11) $y_n = y_{n-2} + kC^{n-2}$

where

(12) $y_n = \dfrac{Q_b(3,n)}{\left(f_{22}^0\right)^{n/2}}, \quad k = \left(\dfrac{f_{22}^1 f_{11}^0}{f_{22}^0}\right), \quad C = \dfrac{f_{12}^0}{\sqrt{f_{22}^0}}$

The solution to equation (11) is a summation over the term kC^{2i}. Choosing a lower bound of 1 for the summation and putting in an arbitrary additive constant, the series can be summed as follows:

(13)

$$y_n = k \sum_{i=1}^{(n-2)/2} C^{2i} + L = k\,\frac{C^2 - C^n}{1 - C^2} + L_2, \quad n \text{ even}$$

$$y_n = k \sum_{i=1}^{(n-3)/2} C^{2i+1} + L = k\,\frac{C^3 - C^n}{1 - C^2} + L_3, \quad n \text{ odd}$$

It is clear from evaluating these equations for y_2 and y_3 that $L_2 = y_2$ and $L_3 = y_3$. Retrieving the original variables from equation (12) gives, finally:

(14) $Q_b(3,n) = \begin{cases} \left(f_{22}^0\right)^{n/2} + f_{22}^1 f_{11}^0 \dfrac{\left(f_{22}^0\right)^{n/2} - \left(f_{12}^0\right)^n}{f_{22}^0 - \left(f_{12}^0\right)^2}, & n \text{ even} \\[2em] \left(f_{22}^0\right)^{(n-1)/2}\left(f_{21}^0 + f_{21}^1 f_{11}^0\right) + f_{22}^1 f_{12}^0 f_{11}^0 \dfrac{\left(f_{22}^0\right)^{(n-1)/2} - \left(f_{12}^0\right)^{n-1}}{f_{22}^0 - \left(f_{12}^0\right)^2}, & n \text{ odd} \end{cases}$

where use was made of $Q_b(3, 2) = f^0{}_{22} + f^1{}_{22}f^0{}_{11}$ and $Q_b(3, 3) = f^0{}_{22}\,(f^0{}_{21} + f^1{}_{21}f^0{}_{11}) + f^1{}_{22}f^0{}_{12}f^0{}_{11}$.

The solution for $n = 2$ and arbitrary m, although not technically in the boundary region, can be obtained in a similar fashion.

(c)   **Combined Solution**

The general solution for the probability of an attacker win, valid for all points in the A-D plane, can be stitched together from the free-space solution and the boundary region solutions by multiplying the probability of reaching each boundary region point by the probability of victory from that point, and summing. The "boundary" points must be chosen such that they are reachable purely through 3 vs. 2 rolls so that the free-space solution can be correctly applied.

Care must be taken, however, to make sure that all paths to victory are counted once and once only. For example, consider the point marked "1" on the edge of the free-space region in Figure 2. There are three possible ways to get there with one roll. One of these ways, however, is from a point that is also on the edge of

the free-space region (marked point "2"). If we were to multiply the probability of reaching point 1 times the probability of victory from point 1, and then do the same thing for point 2, both of those terms would include the path that joins points 1 and 2. Summing all such terms therefore would lead to double-counting of many of the paths to victory.

In order to prevent this, we must construct our solution out of terms comprised of the following factors:

- $P^*(M, N, m, n)$ = probability of *first* reaching the boundary region at point (m, n) from the starting point (M, N) – excludes paths reaching (m, n) from another boundary point.
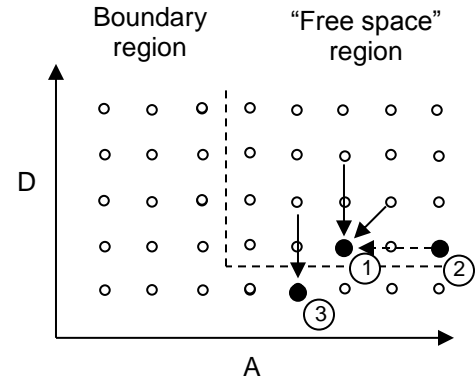- $Q_b(m, n)$ = probability of an attacker victory from starting point (m, n).

Figure 2. **Ways of reaching the boundary region.**

P^* is easily constructed from our existing free-space solution by counting only the paths indicated by the solid arrows in Figure 2. For points on the edge of the free-space region, the equation is:

(15)
$$P^*(M,N,m,2) = f_{32}^1 P_{m+1,3} + f_{32}^0 P_{m,4}$$
$$P^*(M,N,4,n) = f_{32}^1 P_{5,n+1} + f_{32}^2 P_{6,n}$$

We also must look at points one row closer to the axes, such as point "3" in Figure 2, since they can also be the target of a first-time entry into the boundary region. For such points the only path of interest is the one which "leaps over" the edge of the free-space region, either by losing two defender armies (if $n = 3$) or by losing two attacker armies (if $m = 5$):

(16)
$$P^*(M,N,m,1) = f_{32}^0 P_{m,3}$$
$$P^*(M,N,3,n) = f_{32}^2 P_{5,n}$$

The general solution, $Q(M, N)$, for the probability of an attacker victory from an arbitrary starting point (M, N) is:

(17) $$Q(M,N) = \sum \sum_{\substack{m,n \text{ where} \\ m=3,4 \& n\geq2 \\ n=1,2 \& m\geq3}} P^*(M,N,m,n)Q_b(m,n)$$

A javascript implementation of equations (5) – (8), (14), and (17) is available at http://www.recreationalmath.com/risk/riskprob.htm.

Attacking Multiple Countries Sequentially