BINOMIAL BOWLING

Patrick T. Brown and David K. Neal

The game of bowling provides an inherent mathematical structure that can be analyzed under various probabilistic scenarios. Although there is not much randomness involved for skilled bowlers, novices may feel that they are simply knocking down pins at random. In fact, suppose we play by knocking down anywhere from 0 to 10 pins at random, with each number of pins being equally likely, and then roll a second ball that also knocks down pins at random from the remainder. Then the average score over ten frames would be about 91.4127 [2], which is not a bad score for beginners. But this average differs from the mean score of all possible games when the scores are weighted according to their frequencies. In [1], it is shown that there are approximately 5.7 billion billion possible ways to bowl a game and that the average score is about 80.

In this article, we shall consider another scenario in which the pins are knocked down according to binomial distributions. On the first ball of each frame, each pin has equal probability p_1 of being knocked down and pins fall independently of each other. On the second ball of each frame, each remaining pin has equal probability p_2 of being knocked down, and again pins fall independently of each other. Spares and strikes are accounted for according to the regular rules of bowling. Under these conditions, we shall derive the average bowling score for a game with N pins, and find the correlation between the first roll and second roll in each frame. Moreover, we shall see that the second roll and the resulting sum of the two rolls still follow binomial distributions over the range of N pins.

The Binomial Distribution. A binomial distribution $X \sim b(n, p)$ counts the number of successes in n independent attempts, where p is the probability of success on each attempt. For $0 \le k \le n$, the probability of exactly k successes is given by

$$P\left(X = k\right) = \binom{n}{k} p^k \left(1 - p\right)^{n-k},$$

the average number of successes is

$$E[X] = \sum_{k=0}^{n} kP\left(X = k\right) = np,$$

and the variance is

$$Var(X) = np(1-p).$$

Because

$$Var(X) = E[X^2] - (E[X])^2,$$

then

$$E[X^2] = (E[X])^2 + Var(X) = n^2p^2 + np(1-p).$$

We shall use these common results throughout, as well as the binomial expansion theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

The Rules of Binomial Bowling. There are M frames and two rolls per frame $\{X(i,1), X(i,2)\}$, where $X(i,1)$ is from 0 to N according to a $b(N, p_1)$ distribution and $X(i,2)$ follows a $b(N - X(i,1), p_2)$ distribution for $i = 1, 2, \ldots, M, M+1$. If $X(i,1) = N$, then $X(i,2)$ must equal 0. Initially, the score $f(i)$ for the ith frame is the sum of these two rolls. But if $X(i,1)+X(i,2) = N$ and $X(i,1) < N$, which is a spare, then the next roll $X(i+1,1)$ is added to $f(i)$. This roll $X(i+1,1)$ again is chosen according to a $b(N, p_1)$ distribution and is independent of any previous roll. If a spare occurs in the Mth frame, then $X(M+1,1)$ is defined, but $X(M+1,2)$ is left undefined.

A strike occurs if $X(i,1) = N$. If so, then the next two rolls are added to $f(i)$. Because $X(i,2)$ will equal 0, the next roll is actually $X(i+1,1)$ and is independently performed on a new set of pins. If $X(i+1,1) = N$ also, then the following roll is $X(i+2,1)$ and is also on a new set of pins. But if $X(i+1,1) < N$, then the second roll of the additional two rolls is $X(i+1,2)$ and is chosen according to a $b(N - X(i+1,1), p_2)$ distribution. If a strike occurs in the Mth frame, then $X(M+1,1)$ and either $X(M+1,2)$ or $X(M+2,1)$ are defined.

Therefore, the score in the ith frame is given by

$$\begin{aligned} f(i) = {} & (X(i,1) + X(i,2)) + X(i+1,1) \times 1_{\{X(i,1)+X(i,2)=N \cap X(i,1)<N\}} \\ & + (N + X(i+2,1)) \times 1_{\{X(i,1)=N\}} \times 1_{\{X(i+1,1)=N\}} \\ & + (X(i+1,1) + X(i+1,2)) \times 1_{\{X(i,1)=N\}} \times 1_{\{X(i+1,1)<N\}}. \end{aligned}$$

The final score is

$$f = \sum_{i=1}^{M} f(i);$$

and because the average of a sum is the sum of the averages, the average final score will be $M \times E[f(i)]$.

The Average Sum of Two Rolls. So suppose that we bowl M frames on N pins that are knocked down according to the described binomial distributions $\{X(i,1), X(i,2)\}$. Then on the first roll of each frame, the average number of pins knocked down is $E[X(i,1)] = Np_1$ for $i = 1, 2, \ldots, M$.

The range of the second roll is dependent on the first roll. For instance, if a 6 is rolled on the first ball, then the score on the second roll is chosen from the integers 0 to $N-6$ according to a $b(N-6, p_2)$ distribution. Thus, for $i = 1, 2, \ldots, M$, we have

$$X(i,2) \sim b(N - X(i,1), p_2) = \sum_{k=0}^{N} b(k, p_2) \times 1_{\{X(i,1)=N-k\}}.$$

So to find the average of the second roll, we merely weight the averages of the $b(k, p_2)$ distributions by the probabilities of $X(i,1)$ equaling $N-k$. The average of $X(i,2)$ is then

$$\begin{aligned}
E[X(i,2)] &= \sum_{k=0}^{N} E[b(k,p_2)] \times P(X(i,1) = N-k) \\
&= \sum_{k=0}^{N} k \times p_2 \times \binom{N}{N-k} p_1^{N-k}(1-p_1)^k \\
&= p_2 \sum_{k=0}^{N} k \times \binom{N}{k} p_1^{N-k}(1-p_1)^k \\
&= p_2 \times E[b(N, 1-p_1)] = p_2 \times N \times (1-p_1) \\
&= N(p_2 - p_1 p_2) = Np_2(1-p_1). \qquad (1)
\end{aligned}$$

So without taking into account strikes and spares, the average score per frame is

$$E[X(i,1)] + E[X(i,2)] = N(p_1 + p_2 - p_1 p_2). \tag{2}$$

We note that when $N = 10$ and $p_1 = p_2 = 0.5$, then this average is 7.5, which is the same value obtained when knocking down pins at random according to discrete uniform distributions [2]. From Equation (1), we also can observe the obvious behavior of $E[X(i,2)]$ as p_1 increases to 1 or decreases to 0.

Accounting for Spares. If a spare is rolled in a frame, then the next roll is added to the score. Because the next roll is on a new set of pins, the average score is Np_1. But the probability of a spare must be found in order to determine the average amount added to each frame's score. A spare on N pins occurs if $X(i,1) + X(i,2) = N$ but $X(i,1) < N$. For example, if $X(i,1) = k$, then $X(i,2)$ must equal $N - k$, which occurs with probability p_2^{N-k} because the second roll is only on $N - k$ pins. In general, the probability of rolling a spare is

$$\begin{aligned}
&P\left(X(i,1) + X(i,2) = N \cap X(i,1) < N\right) \\
&= \sum_{k=0}^{N-1} P(X(i,1) = k) \times P(X(i,2) = N - k) \\
&= \sum_{k=0}^{N-1} \binom{N}{k} p_1^k (1 - p_1)^{N-k} \times p_2^{N-k} \\
&= \sum_{k=0}^{N-1} \binom{N}{k} p_1^k (p_2 - p_1 p_2)^{N-k} \\
&= (p_1 + p_2 - p_1 p_2)^N - p_1^N. \qquad (3)
\end{aligned}$$

Thus, the average addition per frame to account for spares is given by

$$Np_1 \times \left((p_1 + p_2 - p_1 p_2)^N - p_1^N\right).$$

Accounting for Strikes. If a strike is rolled in a frame, then the next two rolls are added to the frame's score. There are two cases in this situation. Either the next roll after the strike is also a strike, or the next roll is not a strike. In the first case, the average addition to the original frame is N, from the second strike, plus the average score of Np_1 from the next roll on a new set of pins. This first case occurs only in the event of two strikes in a row which has probability $p_1^N \times p_1^N$. Thus, the average addition to a frame in this case is $N(1+p_1)p_1^{2N}$.

But if a strike is rolled, with probability p_1^N, and the next roll is not another strike, then the average addition to the frame's score is given by the average of the next two rolls given that the first of these rolls is not N. Equivalently, it is the unconditional average $N(p_1+p_2-p_1p_2)$ of the next two rolls minus the conditional average of the next two rolls given that the first roll is N (and the second roll is automatically a 0). So the average addition to a frame in this situation is

$$
\begin{aligned}
&P(X(i,1)=N) \times E\left[X(i+1,1)+X(i+1,2) \mid X(i+1,1)<N\right] \\
&= p_1^N \times \big(E\left[X(i+1,1)+X(i+1,2)\right] \\
&\qquad - E\left[X(i+1,1)+X(i+1,2) \mid X(i+1,1)=N\right]\big) \\
&= p_1^N \times (N(p_1+p_2-p_1p_2) - Np_1^N).
\end{aligned}
$$

The Average Score. Finally, the average score per frame in N-pin binomial bowling is

$$
\begin{aligned}
&E[f(i)] = N(p_1+p_2-p_1p_2) + Np_1\big((p_1+p_2-p_1p_2)^N - p_1^N\big) \\
&+ N(1+p_1)p_1^{2N} + p_1^N\big(N(p_1+p_2-p_1p_2) - Np_1^N\big) \\
&= N(p_1+p_2-p_1p_2) + Np_1(p_1+p_2-p_1p_2)^N + Np_1^N(p_1^{N+1}+p_2-p_1p_2),
\end{aligned}
$$

and the average game score over M frames is

$$
E[f] = M\big(N(p_1+p_2-p_1p_2)+Np_1(p_1+p_2-p_1p_2)^N+Np_1^N(p_1^{N+1}+p_2-p_1p_2)\big).
$$

In particular, for a ten-frame game on 10 pins with $p_1 = p_2 = 0.5$, the average score is about 77.84. Also, $p_1 = 1$ produces a perfect score of $3MN$, and $p_1 = 0$ with $p_2 = 1$ produces a score of MN.

Other Binomial Distributions. By assumption, the first roll on each frame follows a binomial distribution: $X(i,1) \sim b(N, p_1)$. Once the value of $X(i,1)$ is determined, then $X(i,2)$ is defined to be binomial over the *remaining* pins. However, we assert that $X(i,2)$ also follows a binomial distribution over the range of all N pins. From its average of $N(p_2 - p_1p_2)$

in Equation (1), we can conjecture that $X(i,2) \sim b(N, p_2 - p_1 p_2)$. Indeed for $0 \leq k \leq N$, we have

$$P(X(i,2) = k) = \sum_{j=0}^{N-k} P(X(i,1) = j) \times P\left(b(N-j, p_2) = k\right)$$

$$= \sum_{j=0}^{N-k} \binom{N}{j} p_1^j (1-p_1)^{N-j} \times \binom{N-j}{k} p_2^k (1-p_2)^{N-j-k}$$

$$= \frac{p_2^k}{k!} \sum_{j=0}^{N-k} \frac{N!}{j!(N-j-k)!} p_1^j (1-p_1)^{N-j} (1-p_2)^{N-j-k}$$

$$= \frac{P(N,k) p_2^k (1-p_1)^k}{k!} \sum_{j=0}^{N-k} \frac{(N-k)!}{j!(N-j-k)!} p_1^j \big((1-p_1)(1-p_2)\big)^{N-j-k}$$

$$= \binom{N}{k} (p_2 - p_1 p_2)^k \times \big(p_1 + (1-p_1)(1-p_2)\big)^{N-k}$$

$$= \binom{N}{k} (p_2 - p_1 p_2)^k (1 - p_2 + p_1 p_2)^{N-k},$$

which is the probability distribution function of a binomial distribution with probability parameter $p_2 - p_1 p_2$. Hence, $X(i,2) \sim b(N, p_2 - p_1 p_2)$ and, as we have seen, its average must be $N(p_2 - p_1 p_2)$.

We also assert that the sum $S(i) = X(i,1) + X(i,2)$ follows a binomial distribution over the range of all N pins. In this case,

$$P(S(i) = k) = \sum_{j=0}^{k} P(X(i,1) = j) \times P\left(b(N-j, p_2) = k - j\right)$$

$$= \sum_{j=0}^{k} \binom{N}{j} p_1^j (1-p_1)^{N-j} \times \binom{N-j}{k-j} p_2^{k-j} (1-p_2)^{N-k}$$

$$= (1-p_2)^{N-k} \sum_{j=0}^{k} \frac{N!}{j!(k-j)!(N-k)!} p_1^j (1-p_1)^{N-k} (1-p_1)^{k-j} p_2^{k-j}$$

$$= \frac{P(N,k)(1-p_2)^{N-k}(1-p_1)^{N-k}}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} p_1^j (p_2 - p_1p_2)^{k-j}$$

$$= \binom{N}{k} (1 - p_1 - p_2 + p_1p_2)^{N-k} (p_1 + p_2 - p_1p_2)^k,$$

which is the probability distribution function of a binomial distribution with probability parameter $p_1 + p_2 - p_1p_2$. Thus, on each frame, the sum of the two rolls $X(i,1) + X(i,2)$ follows a $b(N, p_1 + p_2 - p_1p_2)$ distribution. As we have seen in Equation (2), its average must be $N(p_1 + p_2 - p_1p_2)$.

Finally, if we do not account for strikes and spares, then the scores $S(i)$ on each frame are independent of each other. It is known that the sum of independent binomial distributions with a common probability parameter remains binomial over the sum of the ranges. Therefore, if we do not account for strikes and spares, then the overall score over M frames follows a $b(MN, p_1 + p_2 - p_1p_2)$ distribution.

Correlation Between the Rolls. On each frame, the second roll is clearly dependent on the first roll, and its value decreases as the value of the first roll increases. Hence, there should be a negative correlation between these values. To find this correlation, we first compute the average product as follows:

$$E[X(i,1) \times X(i,2)] = E\left[\sum_{k=0}^{N} k \times b(N-k, p_2) \times 1_{\{X(i,1)=k\}}\right]$$

$$= \sum_{k=0}^{N} k \times (N-k)p_2 \times P(X(i,1) = k)$$

$$= Np_2 \sum_{k=0}^{N} k \times P(X(i,1) = k) - p_2 \sum_{k=0}^{N} k^2 \times P(X(i,1) = k)$$

$$= Np_2 \times E[X(i,1)] - p_2 \times E[X(i,1)^2]$$

$$= Np_2 \times Np_1 - p_2 \left((Np_1)^2 + Np_1(1-p_1)\right)$$

$$= N^2 p_1p_2(1-p_1) - Np_1p_2(1-p_1)$$

$$= (N^2 - N)p_1p_2(1-p_1).$$

We now can obtain the correlation ρ between the first and second rolls in a frame. For $0 < p_1 < 1$ and $0 < p_2 \leq 1$, we have

$$\begin{aligned}
\rho &= \frac{E[X(i,1) \times X(i,2)] - E[X(i,1)] \times E[X(i,2)]}{\sqrt{Var(X(i,1))}\sqrt{Var(X(i,2))}} \\
&= \frac{(N^2 - N)p_1p_2(1-p_1) - Np_1 \times Np_2(1-p_1)}{\sqrt{Np_1(1-p_1)}\sqrt{Np_2(1-p_1)(1-p_2+p_1p_2)}} \\
&= \frac{-Np_1p_2(1-p_1)}{N(1-p_1)\sqrt{p_1}\sqrt{p_2}\sqrt{1-p_2+p_1p_2}} \\
&= \frac{-\sqrt{p_1p_2}}{\sqrt{1-p_2+p_1p_2}}.
\end{aligned}$$

We note that when $p_2 = 1$, then $X(i,2)$ is precisely the linear function $N - X(i,1)$, which yields $\rho = -1$.

Exercises. Here are some interesting calculus exercises based upon the derived probabilities and averages.

1. For a fixed p_2 in Equation (3), find the value of p_1 that maximizes the probability of obtaining a spare.
2. Find the value of p_1 that maximizes the probability of obtaining a strike followed by a non-strike, or that maximizes the probability of obtaining two strikes in a row followed by a non-strike.
3. For a fixed value of p_1+p_2, find the values of the probability parameters that maximize (or minimize) the average score.

<u>*References*</u>

1. C. N. Cooper and R. E. Kennedy, "Is the Mean Bowling Score Awful?," *Journal of Recreational Mathematics,* 18 (1985-86), 205–208.
2. D. K. Neal, "Random Bowling," *Journal of Recreational Mathematics,* Vol. 32 (2), (2004), 125–129

Mathematics Subject Classification (2000): 60E05, 60K40

Patrick T. Brown
Department of Mathematics
Western Kentucky University
Bowling Green, KY 42101
email: ptb1mus@aol.com

David K. Neal
Department of Mathematics
Western Kentucky University
Bowling Green, KY 42101
email: david.neal@wku.edu