Uniform distributions of Discrete Random Variable Sums: An Algebraic Approach

Bill Forrest, Santa Clara University

Oregon State University REU Program

Summer 1991

Introduction

A common question in many introductory courses in undergraduate probability is the following: "Is it possible to load a pair of dice so that the probability of occurrence of every sum from 2 through 12 is equal?" If we consider the dice as discrete, independent random variables, the question can be rephrased as "Is the sum of two independent random variables assuming integer values from 1 through 6 ever uniform?" The answer (as we will show later on in this paper) is no.

This question does, however, raise the further question of under which conditions the sum of finitely many discrete, independent random variables is uniform, and under which it is non-uniform. In addressing these problems, some work in the past (including a solution to the question above) has used the method of moment-generating functions. I have taken a more combinatorial approach to the question, relying on algebraic manipulation and chains of inequality. Proofs by contradiction are used freely.

1. Conventions:
   a. Throughout the paper, all random variables are discrete and independent.
   b. We say that a random variable R assumes a value v if R assumes v with nonzero probability.
   c. If a random variable B assumes a value b, we say that the probability that B=b, normally written as P(B=b), will be denoted simply by P(b).
   d. All random variables assume two or more values.

Definition: The sum on n random variables D_1, D_2,..., D_n is the random variable $W=D_1+D_2+\ldots+D_n$ whose values are all the distinct real numbers assumed by adding together all possible combinations of values of D_1, D_2,..., D_n.

example: Let A and B be two uniform random variables assuming values 0 and 1. Then the sum of A and B, denoted by (A+B), assumes values 0, 1, and 2. Furthermore,

$P([A+B]=0) = P(A=0)P(B=0) = (1/4)$.
$P([A+B]=1) = P(A=1)P(B=0) + P(A=0)P(B=1) = (1/2)$
$P([A+B]=2) = P(A=1)P(B=1) = (1/4)$.

When studying the uniformity of convolutions of random variables, we discover that the following three conditions form a "mutual triangle", meaning that the truth of any two of them implies the third.

Theorem #1: Given two random variables A and B, where A assumes values $a_1, a_2, \ldots, a_p$, and B assumes values $b_1, b_2, \ldots, b_p$, if any two of the following three conditions hold, then the third one holds true also.

I) (A+B) is uniform.
II) A is uniform and B is uniform.
III) Each sum of (A+B) has a unique representation.

Proof:
Case #1: Given (A+B) is uniform
A is uniform and B is uniform

Prove each sum of (A+B) has a unique representation.

Proof: Since A and B are both uniform, we let "p" denote the probability of any given sum appearing on A, and "q" the probability of any given sum appearing on B. Now consider the minimal value on (A+B). It can be attained only by taking the minimal value on A and adding to it the minimal value on B. If we let the minimal sum be denoted by a_1+b_1, we see that $P(a_1+b_1) = P(a_1)P(b_1) = (pq)$. Since (A+B) is uniform, this means that the probability of any sum appearing on (A+B) is (pq).

Now, suppose a sum "N" of (A+B) does not have a unique representation. Then, we have $N=a_i+b_{j+1}=a_{i+1}+b_j$ for some values a_i, a_{i+1} of A and b_j, b_{j+1} of B. But then $P(A+B=N) >= P(a_i)P(b_{j+1}) + P(a_{i+1})P(b_j) = pq+pq= 2pq$, but this contradicts the fact that $P(A+B=N) = pq$, since (A+B) is uniform. Therefore, no such sum N exists, so every sum of (A+B) has a unique representation.

Thus, I. and II imply III.

Case #2: Given (A+B) is uniform
each sum of (A+B) has a unique representation,

Prove A is uniform and B is uniform.

Proof: Since each sum of (A+B) has a unique representation, each combination of a value of A and a value of B represents a different sum. Since each of these sums occurs with the same probability, we have that

$$P(b_1)P(a_1)=P(b_1)P(a_2)=....=P(b_1)P(a_p).$$

Since $P(b_1)>0$, we can divide through by it to get

$$P(a_1) = P(a_2) =.....= P(a_p), \text{ so A is uniform.}$$

Similarly, $P(a_1)P(b_1) = P(a_1)P(b_2) =....= P(a_1)P(b_k)$, so division gives

$$P(b_1) = P(b_2) =....= P(b_k), \text{ so B is uniform.}$$

Thus, I. and III. imply II.

Case #3: Given A is uniform and B is uniform, and each sum of (A+B) has a unique representation,

Prove (A+B) is uniform.

Proof: Since each sum of (A+B) has a different representation, we know that each sum has a representation as a_i+b_j. Since A is uniform, let the probability of any given sum appearing be p. Similarly, since B is uniform, let the probability of any given sum appearing be q. Then, the probability of any given sum appearing is (pq). However, since each sum of (A+B) has a unique representation, we know that each value of (A+B) appears with probability (pq). Thus, (A+B) is uniform, and we have

II. and III. imply I.

And we are done.

corollary: Given n random variables $D_1, D_2, \ldots, D_n$, and given the following three statements, the truth of any two of the statements implies the truth of the third:

I) $(D_1+D_2+\ldots+D_n)$ is uniform.
II) D_i is uniform for $1 <= i <= n$
III) Each sum of $(D_1+\ldots+D_n)$ has a unique representation.

proof: We use induction. We know the statements hold for n=2. Suppose they hold for n random variables $D_1, D_2, \ldots, D_n$. Then let D_{n+1} be another random variable. We define $W=D_1+D_2+\ldots+D_n$. Because W and D_{n+1} are both random variables, and because we know the statements hold for the sum of 2 random variables. We know that $W+D_{n+1}$ satisfies the statements. But $W+D_{n+1} = D_1+D_2+\ldots+D_n+D_{n+1}$, so the statements hold for the sum of n+1 random variables if they hold for the sum of n random variables, and we are done.

Furthermore, under certain restrictions we can show that (A+B) is uniform if and only if A is uniform, B is uniform, and every sum of (A+B) can be attained in a unique way by combining values of A and B (i.e.- I. if and only if II. and III.).

Theorem #2: Let A and B be random variables, where A assumes "p" values and B assumes 2 values. Then (A+B) is uniform if and only if both A and B are uniform and every value of (A+B) has a unique representation.

Proof:

case #1: Given A and B are uniform and every value of (A+B) has a unique representaion, it follows immediately from case #3 of theorem #1 above that (A+B) is uniform.

case #2: Given (A+B) is uniform,
Show both A and B are uniform and that every sum of (A+B) has a unique representation.

proof: Assume that every sum of (A+B) does not have a unique representation. Then Let N_1 denote the smallest value of (A+B) which can be formed in 2 ways. Since B assumes only two values, denoted by b_1 and b_2, N_1 must have the form $N_1 = a_i+b_1 = a_j+b_2$. If we assume $b_1<b_2$, then we must have $a_i>a_j$. Then, note that $a_j+b_1 < N_1$, and thus has a unique representation. Since (A+B) is uniform, it follows that

$$P(a_j)P(b_1) = P(a_i)P(b_1) + P(a_j)P(b_2).$$

Since all probabilities are assumed to be nonzero, this implies that

$P(a_j)P(b_1) > P(a_j)P(b_2)$, so division by $P(a_j)$ gives

$$P(b_1) > P(b_2).$$

Now, let N_2 denote the largest value of (A+B) which can be formed in 2 ways. Again, this must have the form $N_2 = a_n+b_1 = a_m+b_2$, so following the same procedure as above, we must have $a_n > a_m$. Then, note that $b_2+a_n > N_2$, so (b_2+a_n) must have a unique representation. It follows from the uniformity of (A+B) that

$$P(b_2)P(a_n) = P(b_2)P(a_m) + P(b_1)P(a_n).$$

Again, all probabilities are positive, so we must have

$P(b_2)P(a_n) > P(b_1)P(a_n)$. Division by $P(a_n)$ gives

$$P(b_2) > P(b_1).$$

Thus, the assumption that (A+B) has values without unique representation leads to the contradiction that $P(b_1) < P(b_2)$ and $P(b_1) > P(b_2)$. Since this is impossible, we know that every value of (A+B) must have a unique representation. Then, since (A+B) is uniform and every value of (A+B) has a unique representation, we know from case #2 of theorem #1 that both A and B are uniform. Therefore, we have shown that when B assumes only 2 values, (A+B) uniform implies that every sum has a unique representation and that both A and B are uniform, and we are done.

corollary: Given n random variables, if any one of them assumes only 2 values, then the convolution of the n random variables is nonuniform.

proof: Let the random variable assuming 2 random variables be denoted by B. Let the sum of all the other n-1 random variables be denoted by A. Then, by theorem #2 above, (A+B) is non-uniform, so the sum of the n random variables is non-uniform, and we are done.

Definition: A die is a random variable which assumes only consecutive integer values.

Note that the probability distribution for a die need not be uniform. Put simply, the die may be "loaded", making some outcomes more likely than others.

We now present a theorem with rather contrived conditions. Despite this, they are conditions which quite often exist in "practical" examples of random variable convolution.

Theorem #3: Let $D_1, D_2, \ldots, D_n$ denote n random variables. Let s_i denote the minimal value on D_i, and l_i the maximal value on D_i. Let at least two of the random variables (which without loss of generality we call D_1 and D_2) satisfy the following conditions:
1) There exists a real number r_1 so that (s_1+r_1) and (l_2-r_1) are values assumed by D_1 and D_2, respectively.
2) There exists a real number r_2 so that (l_1-r_2) and (s_2+r_2) are values assumed by D_1 and D_2, respectively.

Then the sum of the D_i is non-uniform.

Proof: We use a proof by contradiction. Assume that the convolution of the D_i is uniform.

To simplify notation, we define $W = D_1+D_2+\ldots+D_n$. Let "V" denote the number of values assumed by W. Since W is assumed to be uniform, each value occurs with probability (1/V). Let K denote the minimal value of W and M the maximal value of W. Finally, since all the D_i have distinct greatest and least values, we can define

$$P(s_i)=(1/a_i), \text{ and } P(l_i)=(1/b_i).$$

For the sum $l_1 + (s_2+s_3+\ldots+s_n)$, note that

$$P(W= l_1 + (s_2+s_3+\ldots+s_n)) >= (1/b_1)(1/a_2a_3\ldots a_n)$$

Since (l_1-r_2) and (s_2+r_2) are assumed with nonzero probability by D_1 and D_2, respectively, we know that $(l_1-r_2)+(s_2+r_2)+(s_2+s_3+\ldots+s_n)$ is assumed with nonzero probability. If we let $Y = (l_1-r_2)+(s_2+r_2)+(s_2+s_3+\ldots+s_n) = l_1 + (s_2+s_3+\ldots+s_n)$, and define $P(D_1=(l_1-r_2))=(1/c)$ and $P(D_2=(s_2+r_2))=(1/d)$, then we have

$$(1/b_1)(1/a_2a_3\ldots a_n)+(1/cd)(1/a_3\ldots a_n) <= P(W=Y) = (1/V).$$

Finding a common denominator and adding on the left yields

$$(cd+b_1a_2)/(b_1a_2a_3\ldots a_n)(cd) <= (1/V).$$

Now multiply both sides by $(a_1a_2a_3\ldots a_n)$ to get

$$(a_1cd+b_1a_1a_2)/(b_1cd) <= (1/V)(a_1a_2a_3\ldots a_n).$$

Since $(s_1+s_2+s_3+\ldots+s_n)$ represents the one and only way to add the random variables and get K, the lowest sum, and since $P(W=K)=(1/V)$, we have $P(W=(s_1+s_2+s_3+\ldots+s_n))= 1/(a_1a_2a_3\ldots a_n)$, so

$$1/(a_1a_2a_3\ldots a_n) = 1/V, \text{ so}$$

$$(1/V)(a_1a_2a_3\ldots a_n) = 1, \text{ which means that}$$

$$(a_1cd+b_1a_1a_2)/(b_1cd) <= 1.$$

We separate the expression on the left to get

$$(a_1/b_1) + (a_1a_2)/(cd) <= 1.$$

Since $(a_1a_2)/(cd) > 0$, we have

$$(a_1/b_1) < 1, \text{ so } a_1 < b_1.$$

We now perform an analogous process to obtain the desired contradiction.

Consider the sum $Z = (s_1+l_2+l_3+\ldots+l_n)$.

Since D_1 and D_2 satisfy the conditions set forth in the the hypothesis, we pick a real number r_1 so that (s_1+r_1) and (l_2-r_1) occur on D_1 and D_2, respectively, with nonzero probability.

Define $P(D_1=(s_1+r_1))=(1/g)$, and $P(D_2=(l_2-r_1))=(1/h)$.

Then, since $Z = (s_1+l_2+l_3+\ldots+l_n) = [(s_1+r_1)+(l_2-r_1)+l_3+\ldots+l_n)]$, we have

$[(1/a_1)(1/b_2b_3...b_n)]+[(1/gh)(1/b_3...b_n)] <= P(W=Z)=(1/V)$.

Find a common denominator and add the fractions on the left to get

$(gh+a_1b_2)/(gha_1b_2b_3...b_n) <= (1/V)$

Multiply through by $(1/b_1b_2b_3...b_n)$ to get

$(ghb_1+b_1a_1b_2)/(gha_1) <= (1/V)(b_1b_2b_3...b_n)$.

Now, since $(l_1+l_2+l_3+...+l_n) = M$, the maximal value, and since this is the only way to attain M, we must have

$P(W=(l_1+l_2+l_3+...+l_n)) = (1/V) = 1/(b_1b_2b_3...b_n)$

So $(b_1b_2b_3...b_n)(1/V) = 1$, so

$(ghb_1+b_1a_1b_2)/(gha_1) <= 1$.

Now, separating the fraction on the left, we get

$(b_1/a_1) + (b_1b_2)/(gh) <=1$. Since $(b_1b_2)/(gh) > 0$,

We have $(b_1/a_1) < 1$, so $b_1 < a_1$.

Thus, our assumption that W is uniform has led us to the conclusion that $b_1 < a_1$ and $b_1 > a_1$. This contradiction implies that our assumption was wrong. Therefore, $W = D_1+D_2+...+D_n$ does <u>not</u> have a uniform distribution, and we are done.

corollary: The sum of n dice is not uniform.
proof: Since all the dice assume consecutive integer values, pick two of them D_1 and D_2. Pick $r_1=r_2=1$ and apply the above theorem. It follows directly that the sum is non-uniform.

4. A Uniform Example

A whole class of random variables whose sums are uniform do exist, and fit in nicely with the above theorems.
Let m and q be two natural numbers. Define A to be the uniform random variable which assumes values 0, m, 2m,..., qm. Let B be a uniform random variable which assumes values 1, 2, 3,..., m-1. Then, it is easy to see that (A+B) will take on all integer values from 1 to [(q+1)m]-1. Every value has a unique representation, and the sum itself is uniform.

5. The Extra Lemma

Although I do not use the following lemma in this paper, it is simple but not obvious, and may be useful to anyone else wishing to pursue research in this area.

Lemma: Let A and B be random variables, where a_1 and b_1 are

the minimum values, and a_p and b_k the maximum values on A and B, respectively. If (A+B) is uniform, then $P(a_1)=P(a_p)$ and $P(b_1)=P(b_k)$.

Proof: We show that $P(a_1)=P(a_p)$. The second case is identical.

Let (A+B) assume M values. Since a_1+b_1, the minimal value on (A+B), has a unique representation, we must have $P(a_1+b_1)=P(a_1)P(b_1) = (1/M)$. Now let a_i be any value of A, and we see that $P(a_i+b_1) <= (1/M)$, so

$$P(a_i)P(b_1) <= P(a_1)P(b_1), \text{ so}$$

$$P(a_i) <= P(a_1) \quad \text{for all } i.$$

Similarly, a_p+b_k, the maximum value on (A+B), has a unique representation, and thus $P(a_p)P(b_k) = (1/M)$. Now, let a_i be any value of B. Since $P(a_i+b_k) <= 1/M$, we have

$$P(a_i)P(b_k) <= P(a_p)P(b_k), \text{ so}$$

$$P(a_i) <= P(a_p) \text{ for all } i.$$

Thus, both $P(a_1)$ and $P(a_p)$ are maximal, so they must be equal.

Conclusion: The majority of my research time this summer was spent in a vain attempt to prove a generalization of theorem #2, namely that the sums of random variables (A+B) is uniform if and only if both A and B are uniform and each value of (A+B) has a unique representation as a sum of values of A and B. Although I failed to prove this, I am convinced it is true.

Other open questions related to this paper are:

1) If a sum (A+B) is non-uniform, let a be the largest probability in the distribution and b the smallest. How close can (a/b) come to 1? That is, can we construct convolutions which become arbitrarily close to uniform distributions, or is there a limiting factor in the aforementioned ratio (e.g.- 2)?

2) How can this work be extended to continuous random variables? Is it ever possible to convolute two continuous random variables and get a uniform distribution. If not, why not? If so, under what conditions can we achieve uniformity? I confess that I have not researched this question at all and am thus ignorant of any prior work done in this area.

Finally, I would like to thank Professors Paul Halmos and Robert Bekes of Santa Clara University, and Prof. Paul Cull of Oregon State University for their help on this and related questions, and for their surprising patience.