

On the sum of random variables

December 28, 2013

The purpose of this write up is to write down an organized proof that the expected number of random real numbers $\in [0, 1]$ needed to get a sum greater than 1 is e . I was inspired to try to prove this by reading about this result in the fun book “The Simpsons and their mathematical secrets” by Simon Singh (the statement appears on p. 137).

To make the proof easier to read, I will prove a few independent results first.

Lemma 1. *Let $x \in [0, 1]$ and let $k \in \mathbb{N}$. Then the probability that k random real numbers taken uniformly from $[0, 1]$ sum to a value $\leq x$ is*

$$\frac{x^k}{k!}.$$

Proof. We’re trying to find the probability that $x_1 + x_2 + \dots + x_k \leq x$ when x_i is taken uniformly from $[0, 1]$. The probability is the following:

$$\int_0^x \int_0^{x-x_1} \int_0^{x-x_1-x_2} \dots \int_0^{x-x_1-x_2-\dots-x_{k-1}} 1 \, dx_k dx_{k-1} \dots dx_1.$$

Let’s prove that this integral equals $x^k/k!$ by induction. First, it’s true for $k = 1$ since

$$\int_0^x 1 \, dx_1 = x = \frac{x^1}{1!}.$$

Now let’s assume it’s true for k . Let’s prove that the statement is true for $k + 1$.

$$\begin{aligned} & \int_0^x \int_0^{x-x_1} \int_0^{x-x_1-x_2} \dots \int_0^{x-x_1-x_2-\dots-x_{k-1}-x_k} 1 \, dx_{k+1} dx_k dx_{k-1} \dots dx_1 = \\ & = \int_0^x \left(\int_0^{x-x_1} \int_0^{x-x_1-x_2} \dots \int_0^{x-x_1-x_2-\dots-x_{k-1}-x_k} 1 \, dx_{k+1} dx_k dx_{k-1} \dots dx_1 \right). \end{aligned}$$

So by the induction hypothesis this yields

$$\int_0^x \frac{(x-x_1)^k}{k!} \, dx_1 = -\frac{(x-x_1)^{k+1}}{(k+1)!} \Big|_0^x = \frac{x^{k+1}}{(k+1)!}.$$

□

Theorem 1. *Let $k \in \mathbb{N}$. Let $x_1, x_2, x_3, \dots, x_k$ be randomly chosen real numbers from the interval $[0, 1]$. The probability that $x_1 + \dots + x_k > 1$ while $x_1 + \dots + x_{k-1} \leq 1$ is 0 if $k = 1$ and*

$$\frac{1}{k(k-2)!},$$

if $k \geq 2$.

Proof. It's clearly true for $k = 1$, so let's assume $k \geq 2$. Since $x_1 + x_2 + \dots + x_{k-1} \leq 1$, let $x_1 + x_2 + \dots + x_{k-1} = x$. Then for the sum to surpass 1, $x_k > 1 - x$. The probability that $x_k > 1 - x$ is x . The probability that $x_1 + x_2 + \dots + x_{k-1} = x$ is 0, so what we're going to do is pick a range, so let's say $x_1 + x_2 + \dots + x_{k-1} \in [x, x + \varepsilon]$ for some $\varepsilon > 0$.

The probability that $x_1 + x_2 + \dots + x_{k-1} \in [x, x + \varepsilon]$ is

$$\frac{(x + \varepsilon)^{k-1}}{(k-1)!} - \frac{x^{k-1}}{(k-1)!} = \varepsilon \frac{(k-1)x^{k-2}}{(k-1)!} + O(\varepsilon^2) = \varepsilon \frac{x^{k-2}}{(k-2)!} + O(\varepsilon^2).$$

Therefore the probability that $x_1 + x_2 + \dots + x_k > 1$ while $x_1 + x_2 + \dots + x_{k-1} \leq 1$ is

$$\int_0^1 \frac{x^{k-2}}{(k-2)!} (x) dx = \frac{1}{(k-2)!} \int_0^1 x^{k-1} = \frac{x^k}{k(k-2)!} \Big|_0^1 = \frac{1}{k(k-2)!}.$$

□

Here's a fun side corollary:

Corollary 1.

$$\sum_{k=2}^{\infty} \frac{1}{k(k-2)!} = 1.$$

Proof. Let x_1, x_2, x_3, \dots be randomly chosen real numbers uniformly taken from the interval $[0, 1]$. Let q be the smallest integer such that $x_1 + x_2 + \dots + x_q > 1$. Since $x_i \leq 1$, $q \geq 2$. By the previous theorem, the probability that $q = k$ is

$$\frac{1}{k(k-2)!},$$

since q can take any values from 2 to ∞ and the probabilities add up to 1, we have

$$\sum_{k=2}^{\infty} \frac{1}{k(k-2)!} = 1.$$

□

Now, the main corollary of the theorem:

Corollary 2. *The expected number of random reals uniformly chosen from the interval $[0, 1]$ required for their sum to be greater than 1 is e .*

Proof. Let X be the random variable described in the statement of the corollary. By Theorem 1, the probability that $X = k$ is

$$\frac{1}{k(k-2)!},$$

therefore

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=2}^{\infty} \frac{k}{k(k-2)!} = \sum_{k=0}^{\infty} \frac{1}{k!} = e.$$

□

1 Without using Integrals

What if we want to prove this without using integrals. Here's how one would translate the problem: For a positive integer n , let x_1, x_2, \dots be randomly chosen integers chosen from the set $\{0, 1, 2, \dots, n-1\}$.

It's not that hard to solve this problem with a nice combinatorial formula, but it is a bit harder to recover e as we did above. We will first write down two different formulas that calculate the probability that $x_1 + \dots + x_k \geq n$ while $x_1 + x_2 + \dots + x_{k-1} < n$. Note that to recover e , we should show that as $n \rightarrow \infty$ these formulas go to $\frac{1}{k(k-2)!}$.

For the proofs we're going to need two classic combinatorial results, let me prove them first as lemmas:

Lemma 2. *Let k, m be nonnegative integers. The number of ways $x_1 + x_2 + \dots + x_k = m$ for x_i nonnegative integers is*

$$\binom{m+k-1}{k-1}.$$

Proof. Consider $m+k-1$ slots. On $k-1$ of the slots put a "plus" symbol. The number of empty slots before the first "plus" symbol is x_1 . The number of slots between the first "plus" and the second "plus" is x_2 and so on. Translating these "slots" into k numbers is a bijection, so the number of ways for $x_1 + x_2 + \dots + x_k = m$ is the number of ways of placing those "plus" signs which is precisely

$$\binom{m+k-1}{k-1}.$$

□

Lemma 3 (Chinese Stockings Theorem/ Hockey Stick Theorem).¹

Let r be a nonnegative integer and $n \geq r$ be an integer, then

$$\sum_{k=0}^{n-r} \binom{r+k}{r} = \binom{n+1}{r+1}.$$

Proof. Let's count the number of ways of choosing $r+1$ integers from the set $\{1, 2, 3, \dots, n, n+1\}$ in two different ways.

On the one hand, it is $\binom{n+1}{r+1}$. On the other hand let's count by considering the biggest element in our set. Clearly the biggest element is a number between $r+1$ and n . Let's call this element i . If i is the biggest element, then we have $\binom{i-1}{r}$ ways of choosing the other r elements. Therefore we have that the number of ways of choosing $r+1$ elements from the set $\{1, 2, 3, \dots, n, n+1\}$ is

$$\binom{n+1}{r+1} = \sum_{i=r+1}^{n+1} \binom{i-1}{r} = \sum_{k=0}^{n-r} \binom{k}{r}.$$

□

¹The name of "Chinese Stockings Theorem" or "Hockey Stick Theorem" comes from the shape the summands have in Pascal's Triangle when one selects them.

Now we're ready to prove the main theorem:

Theorem 2. For a positive integer n , let x_1, x_2, \dots be randomly chosen integers chosen from the set $\{0, 1, 2, \dots, n-1\}$. The probability that $x_1 + x_2 + \dots + x_k \geq n$ while $x_1 + x_2 + \dots + x_{k-1} < n$ is

$$\frac{n \binom{n+k-2}{k-1} - \binom{n+k-1}{k}}{n^k}.$$

Proof. First let's find the probability that $x_1 + x_2 + \dots + x_k \geq n$ without restricting $x_1 + x_2 + \dots + x_{k-1}$ to be less than n . We'll do this by finding the probability of the complement, that is we'll find the probability that the sum is less than n .

The number of ways to pick x_1, x_2, \dots, x_k is n^k (since the choices are $0, 1, 2, \dots, n-1$). The probability that the sum $x_1 + x_2 + \dots + x_k$ is m is $\binom{m+k-1}{k-1}/n^k$ (by Lemma 2). Therefore the probability that the sum is less than n is:

$$\sum_{m=0}^{n-1} \frac{\binom{m+k-1}{k-1}}{n^k} = \frac{\binom{n+k-1}{k}}{n^k}.$$

To get the right hand side of the equation we used the Hockey Stick Theorem (Lemma 3).

Combining all of this we have that the probability that the sum is at least n is:

$$1 - \frac{\binom{n+k-1}{k}}{n^k}.$$

Now, to include the bit about $x_1 + x_2 + \dots + x_{k-1} < n$, we take away the times $x_1 + \dots + x_{k-1} \geq n$ from above, so we get that the probability we desire is:

$$1 - \frac{\binom{n+k-1}{k}}{n^k} - \left(1 - \frac{\binom{n+k-2}{k-1}}{n^{k-1}} \right) = \frac{n \binom{n+k-2}{k-1} - \binom{n+k-1}{k}}{n^k}.$$

□

To finish we need to analyze what happens as $n \rightarrow \infty$ when k is fixed. First note that

$$\frac{\binom{n+k-1}{k}}{n^k} = \frac{(n+k-1)(n+k-2)(\dots)(n)}{k!n^k} \rightarrow \frac{1}{k!}.$$

Therefore

$$\frac{n \binom{n+k-2}{k-1}}{n^k} \rightarrow \frac{1}{(k-1)!}.$$

Hence

$$\frac{n \binom{n+k-2}{k-1} - \binom{n+k-1}{k}}{n^k} \rightarrow \frac{1}{(k-1)!} - \frac{1}{k!} = \frac{1}{k(k-2)!}.$$

From this we can see that as $n \rightarrow \infty$, the probability goes to $\frac{1}{k(k-2)!}$, so Corollary 2 follows.