Sum of Uniform Random Variables

Aiden Ament

August 2024

1 Introduction to Problem

I was asked this question in a job interview, and it stumped me at the time. So afterwards, I decided to solve it. Let X_1, X_2, $X_3 \ldots$ be independent, uniformly distributed, random variables on the interval $[0, 1]$. What is the expected number of variables required to reach a cumulative sum greater than 1?

2 Initial Approach

Let's first break this problem down by tackling a simpler question. What is the probability that the sum $X_1 + \cdots + X_n < 1$. First start by defining a different random variable $Y_n := X_1 + X_2 + \cdots + X_n$. We can now express the probability that the sum requires at least n variables as $P(Y_{n-1} < 1 \text{ and } Y_n > 1)$. This is the probability that the sum of the first $n-1$ variables is less than 1 and the sum of the first n variables is greater than 1.

Now we can toy around with simple cases to see how it might motivate a general case. Take $n = 2$ for example. We can calculate $P(Y_1 < 1 \text{ and } Y_2 > 1)$ using an integral.

$$\int_{x_1=0}^{x_1=1} \int_{x_2=1-x_1}^{x_2=1} 1 \; dx_2 \; dx_1 = \int_{x_1=0}^{x_1=1} x_2 \Big|_{x_2=1-x_1}^{x_2=1} dx_1$$

$$= \int_{x_1=0}^{x_1=1} 1 - (1 - x_1) \; dx_1 = \int_{x_1=0}^{x_1=1} x_1 = \frac{1}{2}$$

Its important to note the integral over the entire probability region (an n dimensional unit cube) is 1, so there is no normalization needed. We get the starting integral by integrating first over the region where $X_1 < 1$ and then over the region where $X_1 + X_2 > 1$. Trying with $n = 3$ we can calculate $P(Y_2 < 1 \text{ and } Y_3 > 1)$ by integrating

$$\int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \int_{x_3=1-x_2-x_1}^{x_3=1} 1 \; dx_3 \; dx_2 \; dx_1$$

The first two integrals represent the region where $X_1 + X_2 < 1$ and the final integral represents when $X_1 + X_2 + X_3 > 1$. By calculating the first integral in the expression we get

$$= \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} 1-(1-x_2-x_1) \; dx_2 \; dx_1 = \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} x_1+x_2 \;\; dx_2 \; dx_1$$

Then finishing the evaluation we get

$$= \int_{x_1=0}^{x_1=1} x_1 \, x_2 + \frac{1}{2}x_2^2 \Big|_{x_2=0}^{x_2=1-x_1} dx_1 = \int_{x_1=0}^{x_1=1} x_1(1-x_1) + \frac{1}{2}(1-x_1)^2 \; dx_1$$

$$= \int_{x_1=0}^{x_1=1} \frac{1}{2} - \frac{1}{2}x_1^2 \; dx_1 = \frac{1}{2}x_1 - \frac{1}{6}x_1^3 \Big|_{x_1=0}^{x_1=1} = \frac{2}{6}$$

Now with an intuition for some of the calculations, we can try approaching the general case.

3 General case

To calculate $P(Y_{n-1} < 1 \text{ and } Y_n > 1)$ we can first integrate over a region where $X_1 + X_2 + \cdots + X_{n-1} < 1$ and then have a final integral over the remaining region where $X_1 + \cdots + X_n > 1$. The result is this:

$$\int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1-\cdots-x_{n-2}} \int_{x_n=1-x_1-\cdots-x_{n-1}}^{x_n=1} 1 \; dx_n \, dx_{n-1} \ldots dx_1$$

This expression is long, but we can first begin by evaluating the first integral, and we see a similar pattern to the first two examples.

$$= \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1-\cdots-x_{n-2}} x_n \Big|_{x_n=1-x_1-\cdots-x_{n-1}}^{x_n=1} dA$$

$$= \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1-\cdots-x_{n-2}} 1-(1-x_1-x_2-\cdots-x_n) \; dA$$

$$= \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1-\cdots-x_{n-2}} x_1+x_2+\cdots+x_{n-1} \; dA$$

Now we have made some progress because we can use the linearity of integration to split up this integral into many pieces.

$$= \sum_{k=1}^{n-1} \int_{x_1=0}^{x_1=1} \int_{x_2=0}^{x_2=1-x_1} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1-x_2-\cdots-x_{n-2}} x_k \; dA$$

Let's try and evaluate just one piece. We can further split up one of these pieces for a given x_k by looking at the kth integral.

$$\int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1\cdots-x_{n-2}} x_k \; dA$$

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdots \int_{x_{n-1}=0}^{x_{n-1}=1-x_1\cdots-x_{n-2}} 1 \; dA$$

Now by further evaluating we can notice a pattern.

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdots \int_{x_{n-2}=0}^{x_{n-2}=1-x_1\cdots-x_{n-3}} x_{n-1}\Big|_{x_{n-1}=0}^{x_{n-1}=1-x_1-\cdots-x_{n-2}} dA$$

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdots \int_{x_{n-2}=0}^{x_{n-2}=1-x_1-\cdots-x_{n-3}} 1-x_1-\cdots-x_{n-2} \; dA$$

Then with the u substitution of $u = (1 - x_1 - \cdots - x_{n-2})$ and $du = -dx_{n-2}$ we can get

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdots \int_{x_{n-2}=0}^{x_{n-2}=1-x_1-\cdots-x_{n-3}} -u \; dA$$

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdots \int_{x_{n-3}=0}^{x_{n-3}=1-x_1-\cdots-x_{n-4}} -\frac{1}{2}u^2 \; dA$$

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_{n-3}=0}^{x_{n-3}=1-x_1-\cdots-x_{n-4}} -\frac{1}{2}(1-x_1-\cdots-x_{n-2})^2\Big|_{x_{n-2}=0}^{x_{n-2}=1-x_1-\cdots-x_{n-3}} dA$$

Notice however, when we let $x_{n-2} = 1 - x_1 - \cdots - x_{n-3}$ the entire expression is 0. Therefore when we evaluate the u substitution entirely we get

$$= \int_{x_1=0}^{x_1=1} \cdots \int_{x_{n-3}=0}^{x_{n-3}=1-x_1-\cdots-x_{n-4}} \frac{1}{2}(1 - x_1 - \cdots - x_{n-3})^2 \; dA$$

This can be repeated until we reach the x_k integral. This yields

$$\int_{x_1=0}^{x_1=1} \cdots \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} x_k \cdot \frac{1}{(k-1)!}(1 - x_1 \cdots - x_k)^{k-1} \; dA$$

This can further be evaluated by the same u substitution $u = (1-x_1-\cdots-x_k)$ and $du = -dx_k$. We can also rewrite $x_k = (1-x_1-\cdots-x_{k-1})-u$. This yields

$$\int_{x_1=0}^{x_1=1} \cdots \frac{1}{(k-1)!} \int_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}} u^k - (1 - x_1 - \cdots - x_{k-1})u^{k-1} \; dA$$

This evaluates to

$$\int_{x_1=0}^{x_1=1} \cdots \frac{1}{(k-1)!} \int_{x_{k-1}=0}^{x_{k-1}=1-x_1-\cdots-x_{k-2}} \frac{u^{k+1}}{k+1} - (1 - x_1 - \cdots - x_{k-1})\frac{u^k}{k} \; dA$$

Substituting back in the value of u gets

$$\frac{(1-x_1-\cdots-x_k)^{k+1}}{k+1}-(1-x_1-\cdots-x_{k-1})\frac{(1-x_1-\cdots-x_k)^k}{k}\Bigg|_{x_k=0}^{x_k=1-x_1-\cdots-x_{k-1}}$$

Something very convenient happens again. When $x_k = 1 - x_1 - \cdots - x_{k-1}$ the entire expression evaluates to 0. And thus by letting $x_k = 0$ we get that the result is

$$\begin{aligned}(1-x_1-\cdots-x_{k-1})\frac{(1-x_1-\cdots-x_{k-1})^k}{k}&-\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{k+1}\\ =\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{k}&-\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{k+1}\\ &=\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{k\cdot(k+1)}\end{aligned}$$

Plugging this back into the integral expression we get

$$\begin{aligned}&\int_{x_1=0}^{x_1=1}\cdots\frac{1}{(k-1)!}\int_{x_{k-1}=0}^{x_{k-1}=1-x_1-\cdots-x_{k-2}}\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{k\cdot(k+1)}dA\\ &=\int_{x_1=0}^{x_1=1}\cdots\int_{x_{k-1}=0}^{x_{k-1}=1-x_1-\cdots-x_{k-2}}\frac{(1-x_1-\cdots-x_{k-1})^{k+1}}{(k+1)!}dA\end{aligned}$$

Using the same iterative u substitution technique as we began with, the rest simplifies down to

$$\begin{aligned}&\int_{x_1=0}^{x_1=1}\frac{(1-x_1)^{n-1}}{(n-1)!}dA\\ &=\frac{-(1-x_1)^n}{n!}\Bigg|_{x_1=0}^{x_1=1}=\frac{1}{n!}\end{aligned}$$

Great progress, but this is only a piece of the final result. Recall that our expression for $P(Y_{n-1} < 1 \text{ and } Y_n > 1)$ was

$$\sum_{k=1}^{n-1}\int_{x_1=0}^{x_1=1}\int_{x_2=0}^{x_2=1-x_1}\cdots\int_{x_{n-1}=0}^{x_{n-1}=1-x_1-x_2-\cdots-x_{n-2}}x_k\ dA$$

Note that our result for x_k was actually independent of k, thus our final result is

$$P(Y_{n-1} < 1 \text{ and } Y_n > 1) = \frac{n-1}{n!}$$

Checking for $n = 2$ and $n = 3$ yield $\frac{1}{2}$ and $\frac{2}{6}$ respectively confirming the results we already calculated.

4 Expected Value

Now that we have a precise probability for the sum to require at least n values, we can calculate the expectation.

$$\mathbb{E} = \sum_{n=2}^{\infty} n \cdot \frac{(n-1)}{n!} = \sum_{n=2}^{\infty} \frac{1}{(n-2)!}$$

The sum starts at 2 because there is 0 probability that the first term is greater than 1. Then we can simplify the result by shifting n back by 2 resulting in

$$\mathbb{E} = \sum_{n=0}^{\infty} \frac{1}{n!} = e$$

Which is the Taylor series expansion of the constant e.