If I break a stick of unit length into three random pieces, what's the expected length of the largest piece?

We solve the more general case for the average of the shortest, middle and longest length, when the stick is broken into 3 pieces.

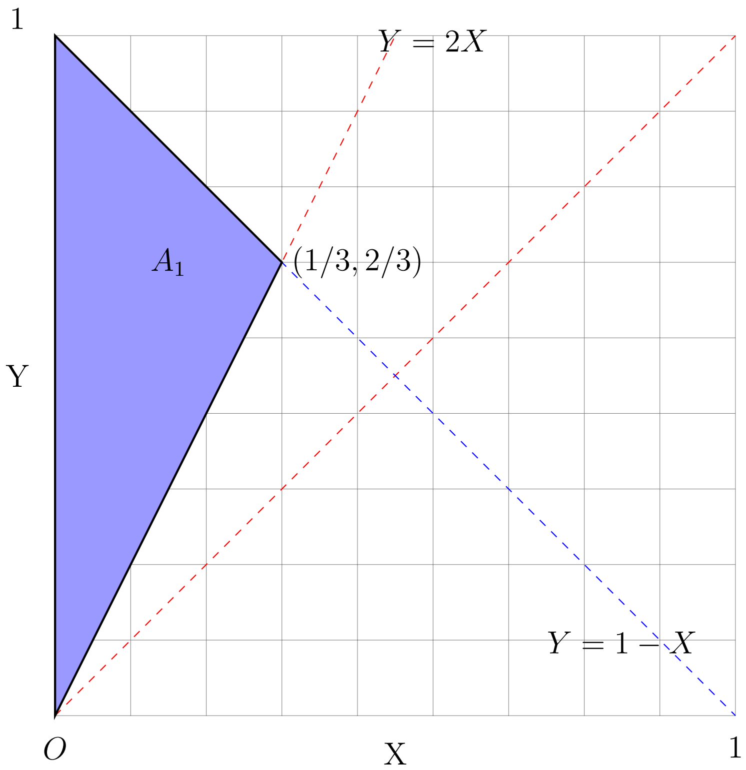
We assume the breaks occur at $x, y \in (0,1)$. Let the ordered pair (X, Y) be a random variable denoting the location of the breaks. Then, (X, Y) is distributed on the unit square with areas are equal to probabilities.

If we break the stick at X and Y, then we can assume that $X < Y$, then the three lengths are

$$\{X, Y - X, 1 - Y\}$$

Without loss of generality, assume X is the smallest length. Then pair-wise comparison of lengths gives

$$X < Y - X \implies Y > 2X, \text{ and,}$$
$$X < 1 - Y \implies Y < 1 - X$$



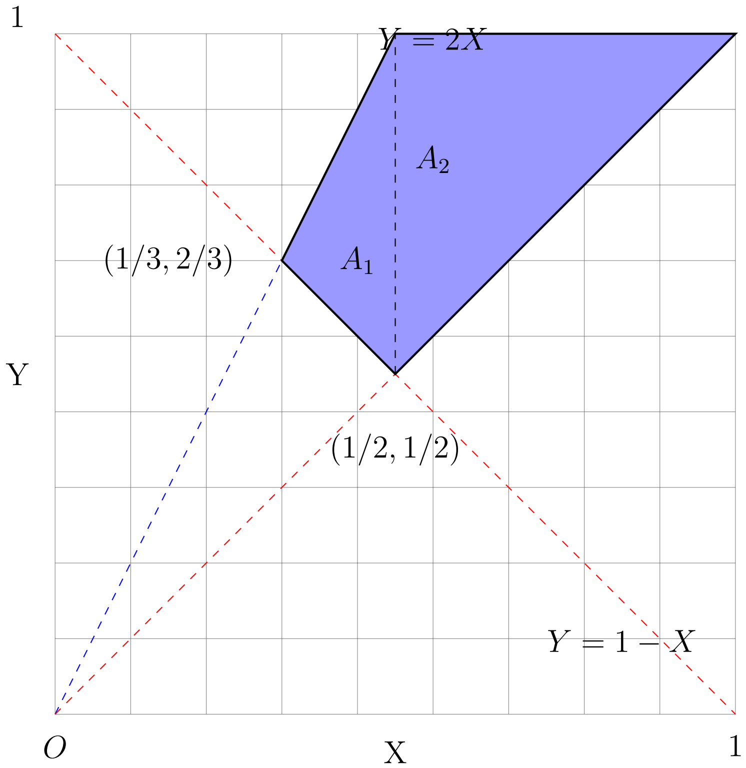
The shaded region A_1 meets all the conditions - to find the $\overline{X}$ we calculate the mean of $x-$coordinate and similarly for $\overline{Y}$ (if it is needed).

$$\overline{X} = (0 + 0 + 1/3)1/3 = 1/9 = 2/18$$

Now assume X is the largest length, then as before we can derive the inequalities

$$X > Y - X \implies Y < 2X, \text{ and,}$$
$$X > 1 - Y \implies Y > 1 - X$$

Recall we still have $X < Y$. So the unit rectangle becomes



In order to find the $\overline{X}$ for the quadrilateral $A_1 \cup A_2$, we need to weight $\overline{X}_1$ and $\overline{X}_2$ for area A_1,A_2 respectively with the areas A_1 and A_2. Mathematically.

$$\overline{X}_T = \frac{A_1\overline{X}_1 + A_2\overline{X}_2}{A_1 + A_2}$$

As before $\overline{X}_1 = (1/2{+}1/3{+}1/2)1/3 = 4/9 = 8/18$ and $\overline{X}_2 = (1/2{+}1{+}1/2)1/3 = 2/3 = 12/18$. For A_1 we can it into two right angle triangles by dropping an altitude to the dotted line (common base) from the opposite vertex to give

$$A_1 = (1/2)b_1(1/2 - 1/3) + (1/2)b_2(1/2 - 1/3) = (1/2)(1/6)(b_1 + b_2) = (1/2)(1/6)(1/2)$$

where $b_1 + b_2$ is the length of the common base between A_1 and A_2 (1/2 in length). Clearly, as A_2 is equilateral then $A_2 = (1/2)(1/2)(1/2)$.
Therefore

$$\begin{aligned} A_1 + A_2 &= (1/2)(1/2)(3/6 + 1/6) = (1/4)(4/6) = 4/24 \\ A_1\overline{X}_1 + A_2\overline{X}_2 &= (1/24)(8/18) + (1/8)(12/18) \\ \implies \frac{A_1\overline{X}_1 + A_2\overline{X}_2}{A_1 + A_2} &= \frac{8/18 + 36/18}{4} = 2/18 + 9/18 = \frac{11}{18} \end{aligned}$$

Since the stick is of unit length - we can now infer the average of the middle length

$$1 - (11/18) - (2/18) = \frac{5}{18}$$

Generalising further, for a stick broken into n pieces, the average lengths satisfy

$$\begin{aligned} \text{smallest:} \quad & \frac{1}{n}\left(\frac{1}{n}\right) \\ \text{(n-1)th smallest:} \quad & \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1}\right) = \frac{1}{n}(H_n - H_{n-2}) \\ & \vdots \\ \text{1st smallest (largest):} \quad & \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + ... + 1\right) = \frac{1}{n}H_n \end{aligned}$$

A general approach to n breaks

Let $0 < L_{(1)} < L_{(2)} < ... < L_{(n)} < 1$ be order statistics from a Dirichlet distribution $D_n(\alpha)$ with n-vector $\alpha = (1, 1, 1, ..., 1)$. By definition,

$$\sum_{i=1}^{n} L_{(i)} = 1.$$

By writing

$$L_{(r)} = L_{(1)} + (L_{(2)} - L_{(1)}) + (L_{(3)} - L_{(2)}) + ... + (L_{(r)} - L_{(r-1)}),$$

for $r \in \{2, 3, ..., n\}$ we can deduce

$$\sum_{i=0}^{n-1} (n-i)(L_{(1+i)} - L_{(i)}) = 1 \qquad (*)$$

where $L_{(0)} = 0$. In order to determine the expectation of the summands,

$$\mathbb{E}[(n-i)(L_{(1+i)} - L_{(i)})], \quad i \in \{0, ..., n-1\}$$

we return to the Dirichlet characterisation by exponential (or gamma) distribution. It is well known that

$$y_1, y_2, ..., y_n \sim Exp(1) \implies \left(\frac{y_1}{\sum_j y_j}, \ldots, \frac{y_n}{\sum_j y_j}\right) \sim D_n((1, \ldots, 1))$$

Considered as a random vector, $\left(\frac{y_1}{\sum_j y_j}, \ldots, \frac{y_n}{\sum_j y_j}\right)$ can be written

$$\Gamma_{n,1}^{-1}(y_1, \ldots, y_n)$$

Importantly, the factor $\Gamma_{n,1}^{-1}$ does not change the ordering of the order statistics. Hence, let

$$0 < Y_{(1)} < Y_{(2)} < \ldots < Y_{(n)}$$

be order statistics for the exponential distribution. Then

$$\mathbb{E}(L_{(1+i)} - L_{(i)}) = \mathbb{E}(Y_{(1+i)} - Y_{(i)})\mathbb{E}(\Gamma_{n,1}^{-1})$$

by independence. Standard results show that

$$\begin{aligned}
Y_{(1)} &\sim Exp(1/n) \\
Y_{(2)} &\sim Y_{(1)} + Exp(1/(n-1)) \\
&\vdots \\
Y_{(r)} &\sim Y_{(r-1)} + Exp(1/(n-r+1))
\end{aligned}$$

Back substituting, we see that

$$\begin{aligned}
Y_{(r)} &\sim Exp(1/n) + Exp(1/(n-1)) + \ldots + Exp(1/(n-r+1)), \\
\mathbb{E}(Y_{(r)}) &= \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-r+1}
\end{aligned}$$

Therefore

$$\mathbb{E}(Y_{(1+i)} - Y_{(i)}) = \frac{1}{n-i}$$

Finally

$$\mathbb{E}[(n-i)(L_{(1+i)} - L_{(i)})] = (n-i)\mathbb{E}(\Gamma_{n,1}^{-1})\mathbb{E}(Y_{(1+i)} - Y_{(i)}) = (n-i)\frac{1}{n(n-i)} = \frac{1}{n}$$

Taking each summand in (*) turn

$$\mathbb{E}(nL_{(1)}) = \frac{1}{n} \implies \mathbb{E}(L_{(1)}) = \frac{1}{n^2}$$

$$\mathbb{E}((n-1)(L_{(2)} - L_{(1)})) = \frac{1}{n} \implies \mathbb{E}(L_{(2)}) = \mathbb{E}(L_{(1)}) + \frac{1}{n}\frac{1}{n-1} = \frac{1}{n^2} + \frac{1}{n}\frac{1}{n-1} = \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1}\right)$$

$$\vdots$$

$$\mathbb{E}((n-i)(L_{(1+i)} - L_{(i)})) = \frac{1}{n} \implies \mathbb{E}(L_{(1+i)}) = \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-i}\right)$$

$$\vdots$$

$$\mathbb{E}(L_{(n)} - L_{(n-1)}) = \frac{1}{n} \implies \mathbb{E}(L_{(n)}) = \frac{1}{n}\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + 1\right) = \frac{1}{n}H_n$$