Sorted Random Numbers. Pick n random numbers $x_1, x_2, ..., x_n$ in the interval $[0, 1]$ with uniform probability. Let $y_1 \leq y_2 \leq \cdots \leq y_n$ be those same numbers sorted in non-decreasing order. For each $k = 1, \ldots, n$, what is the expected value of y_k?

- *Answer:* The expected values $E[y_k]$, $k = 1, \ldots, n$, divide the interval $[0, 1]$ into $n + 1$ subintervals of equal length:

$$E[y_k] = \frac{k}{n+1} \qquad (k = 1, \ldots, n)\,.$$

- *First (Calculus) Solution.*

Given $\lambda \in [0, 1]$, if $y_k = \lambda$, then we have that one of the n given numbers must be λ, $k - 1$ of them must be $\leq \lambda$ (probability $= \lambda^{k-1}$), and the remaining $n-k$ must be $\geq \lambda$ (probability $= (1-\lambda)^{n-k}$). There are n ways to pick the number that equals λ, and $\binom{n-1}{k-1}$ ways to divide $n - 1$ numbers into $k - 1$ less than λ and $n - k$ greater than λ, hence the probability density of $y_k = \lambda$ is:

$$f_k(\lambda) = n\binom{n-1}{k-1}\lambda^{k-1}(1-\lambda)^{n-k} = k\binom{n}{k}\lambda^{k-1}(1-\lambda)^{n-k}\,.$$

Consequently, the expected value of y_k is

$$E[y_k] = \int_0^1 \lambda f_k(\lambda)\,d\lambda = \int_0^1 k\binom{n}{k}\lambda^k(1-\lambda)^{n-k}\,d\lambda\,.$$

The integrand is the coefficient of t^{k-1} in the expansion of

$$\frac{d}{dt}\{t\lambda + (1-\lambda)\}^n\,.$$

Integrating $\{t\lambda + (1-\lambda)\}^n$ respect to λ between 0 and 1 we get:

$$\int_0^1 \{t\lambda + (1-\lambda)\}^n\,d\lambda = \frac{t^{n+1}-1}{(t-1)(n+1)} = \frac{1}{n+1}(1 + t + t^2 + ... + t^n)\,.$$

Differentiating respect to t:

$$\frac{d}{dt}\int_0^1 \{t\lambda + (1-\lambda)\}^n\,d\lambda = \frac{1}{n+1}(1 + 2t + ... + nt^{n-1})\,.$$

The coefficient of t^{k-1} in that expression is the desired expected value:

$$E[y_k] = \frac{k}{n+1} \qquad (k = 1, \ldots, n)\,.$$

- *Second (Non-Calculus) Solution.*

First, note that if x is a random number picked with uniform probability in an interval $[a, b]$ then its expected value is the midpoint of the interval $E[x] = (a + b)/2$. This can be proved directly using the definition of expected value, but it can be justified also by symmetry: if $u = a + b - x$ then u is also a random number with uniform probability in $[a, b]$, so $E[x] = E[u] = E[a + b - x] = a + b - E[x]$, and from here $E[x] = (a + b)/2$.

Second, we note that for $k = 1, \dots, n$, for fixed $y_{k-1} = a$, $y_{k+1} = b$ ($y_0 = 0 \le a \le b \le 1 = y_{n+1}$), y_k is a random number with uniform probability in $[a, b]$, hence $E'[y_k] = (a + b)/2 = (y_{k-1} + y_{k+1})/2$, where E' represents the expected value of y_k subject to the condition $y_{k-1} = a$ and $y_{k+1} = b$. Averaging respect to y_{k-1} and y_{k+1} we get

$$E[y_k] = \frac{1}{2}(E[y_{k-1}] + E[y_{k+1}]) .$$

Hence, $E[y_1], \dots, E[y_n]$ divide the interval $[0, 1]$ into $n + 1$ subintervals of equal length, consequently

$$E[y_k] = \frac{k}{n+1} .$$

- *Third (Simpler Non-Calculus) Solution.*

We can get the result by using the following principle of symmetry: *When n points are dropped at random on an interval, the lengths of the $n+1$ segments in which they divide the interval have identical distributions.*[1] This principle can be justified in the following way. Instead of dropping n points on a unit interval, we drop $n + 1$ points on a circle of length 1; then we map each list of points on the circle to a list of points on the unit interval by cutting at the $(n + 1)$th point and straightening the circle. By circular symmetry it is obvious that the lengths of the $n+1$ intervals in which they divide the circle have identical distributions.

Since the lengths of the $n + 1$ segments have identical distributions, their expected value must be the same, and since the sum of the lengths is 1 each expected value must be $1/(n+1)$. From here the desired result follows.

Miguel A. Lerma - 3/17/2005

[1]See Frederick Mosteller: *Fifty Challenging Problems in Probability with Solutions*, Dover Publications, 1965, pp. 59–60.