Problem of the Week

2020 #14

20160520 Junho Lee

Problem. Let $m_0=n$. For each $i\geq 0$, choose a number x_i in $\{1,\ldots,m_i\}$ uniformly at random and let $m_{i+1}=m_i-x_i$. This gives a random vector $\mathbf{x}=(x_0,x_1,\ldots)$. For each $1\leq k\leq n$, let X_k be the number of occurrences of k in the vector $\mathbf{x}$. For each $1\leq k\leq n$, let Y_k be the number of cycles of length k in a permutation of $\{1,\ldots,n\}$ chosen uniformly at random. Prove that X_k and Y_k have the same distribution.

Proof. For each permutation $\pi\in S_n$, order cycle depending on order of its least element.
Define $c(\pi):=x=\{x_0,x_1,\ldots,x_{l-1}\}$ where x_k is length of k-th cycle.

For any vector $x=\{x_0,\ldots,x_{l-1}\}$, define $m_i(x):=\sum_{k=i}^{l-1}x_k$ and $m_l(x)=0$, $t(x)=\{x_1,\ldots,x_{l-1}\}$.

Clearly, $m_i(x)-m_{i+1}(x)=x_i$ and $m_{i+1}(x)=m_i(t(x))$. By definition of $c(\pi)$, $m_0(c(\pi))=n$.

Claim. Probability of obtaining $x=c(\pi)$ is $P(x)=\prod_{i=0}^{l-1}\frac{1}{m_i(x)}$.

Enough to show there are $n!P(x)$ permutations with $x=c(\pi)$, since the probability of choosing specific permutation π is identically $1/n!$.

Apply induction on l. By definition, 0^{th} cycle always contains 1. Choose remaining x_0-1 elements in the first cycle in order, which gives ${}_{n-1}P_{x_0-1}=\frac{(n-1)!}{(n-x_0)!}$ cases.

If $m=1$, done ($\because$ $x_0=n$, ${}_{n-1}P_{n-1}=(n-1)!=\frac{n!}{n}=\frac{n!}{m_0(x)}=n!P(x)$).

If $m>1$, re-number the remaining of $\{1,\ldots,n\}$ excluding 0^{th} cycle, without changing order, so that it becomes $\{1,\ldots,n-x_0\}$. As we are working with permutation, this does not affect number of cases.
By induction hypothesis, number of cases of choosing the other cycles is $(n-x_0)!P(t(x))$.
Multiplying number of choices of first cycle, we obtain total of $(n-1)!P(t(x))$ possibilities.
Then, $(n-1)!P(t(x))=n!\times\frac{P(t(x))}{m_0(x)}=n!P(x)$. □

Now let $M(x)=\{m_0(x),m_1(x),\ldots,m_l(x)\}$, which is upper bound of x_i in the process of X_k.

Claim. Probability of obtaining **x** from process of X_k is $P(\boldsymbol{x})$.

Consider n to be $m_0(\boldsymbol{x})$. Apply induction on l.

For $l = 1$, n has to be chosen first from $\{1, \ldots, n\}$ and it's done. Probability is $1/n = P(\boldsymbol{x})$.

For $l > 1$, x_0 has to be chosen first from $\{1, \ldots, n\}$, which gives 1/n. After that, it has to choose $t(\boldsymbol{x})$.

Since $m_0(t(\boldsymbol{x})) = m_1(\boldsymbol{x}) = m_0 - x_0$, this is identical to choosing $t(\boldsymbol{x})$ from scratch.

By induction hypothesis, choosing $t(\boldsymbol{x})$ has probability of $P(t(\boldsymbol{x}))$. In total, $\frac{1}{n} \times P(t(x)) = P(x)$. □

Observe that for each vector $\boldsymbol{x}$ with $m_0(\boldsymbol{x})=n$ and $c(\pi)=\boldsymbol{x}$, both X_k and Y_k are occurrences of k in **x**, since the number of k-cycle in π can be counted as occurrence of k in $c(\pi)$.

Obviously, $c(\pi)$ for $\pi \in S_n$ spans entire $\{x : x_k > 0 \, for \; k=1,2,\ldots,m, m_0(x)=n\}$, i.e. both have the same range of **x**. Because $X_k=Y_k$ given $c(\pi)=\boldsymbol{x}$, and probability of choosing $c(\pi)=\boldsymbol{x}$ for both process is the same, X_k and Y_k should have the same distribution.