Exercise 2

Jack sets up to complete an album of N football player sticker cards. To this end, he buys one packet of cards a day. A packet contains k *different* cards, chosen at random from the set of N. Packets are assembled independently, and $N >> k$.

1) Compute the probability $p_{j,1}$ that Jack collects *j* new cards *on* the first day.
2) Compute the probability $p_{j,2}$ that Jack collects *j* new cards *on* the 2nd day. From that, compute the probability $P_{i,2}$ that Jack collects *i* new cards *within* two days.
3) Compute a recursive expression for $p_{j,3}$ (e.g., as a function of either or both $p_{j,2}$, $P_{i,2}$), and generalize it to any number of days *n*.

Solution

1) Since there are no duplicates in a packet, it is obvious that $p_{j,1} = 1_{\{j=k\}}$

2) $p_{j,2}$ is non null for $0 \le j \le k$, and null otherwise. When it is non null, $p_{j,2}$ is the probability that Jack finds j *new* cards and $k-j$ *duplicates* in the packet of *k* cards bought on the 2nd day. By observing that all the results are equally likely, we get that the number of packets of *k* cards that have j *new* cards and $k-j$ *duplicates* is equal to:

$$\binom{N-k}{j} \cdot \binom{k}{k-j}$$

i.e., the number of subsets of $j-k$ cards taken from a set of $N-k$ (i.e., the remaining *new* cards after day 1), times the number of subsets of $2k-j$ taken from a set of k (i.e., the set of cards collected by day 1). The above expression must be divided by the size of the sample space, i.e.:

$$p_{j,2} = \frac{\binom{N-k}{j} \cdot \binom{k}{k-j}}{\binom{N}{k}}, \ 0 \le j \le k \quad (1)$$

Since the number of cards collected on day 1 is constant and equal to *k*, we have $P_{i,2} \ne 0 \Leftrightarrow i \in [k, 2k]$, and $P_{i,2} = p_{i-k,2}$.

3) Again, $p_{j,3}$ is non null for $0 \le j \le k$, and null otherwise, and the same goes for $p_{j,n}$ for any day *n*. Call $p_{x,3|y,2}$ the conditional probability to collect *j new* cards *on* the 3rd, given that you collected *y* *by* the 2nd.

$$p_{j,3|y,2} = \frac{\binom{N-y}{j} \cdot \binom{y}{k-j}}{\binom{N}{k}}, \ 0 \le j \le k, \ k \le y \le 2 \cdot k.$$

Therefore, by the law of total probability, we get:

$$p_{j,3} = \sum_{y=k}^{2k} p_{j,3|y,2} \cdot P_{y,2} = \sum_{y=k}^{2k} \frac{\binom{N-y}{j} \cdot \binom{y}{k-j}}{\binom{N}{k}} \cdot P_{y,2}, \ 0 \le j \le k$$

The above expression can be easily generalized to a generic number of days *n*, by observing that the summation extends from *k* to $(n-1) \cdot k$, i.e.:

$$p_{j,n} = \sum_{y=k}^{(n-1)k} p_{j,n|y,n-1} \cdot P_{y,n-1} = \sum_{y=k}^{(n-1)k} \frac{\binom{N-y}{j} \cdot \binom{y}{k-j}}{\binom{N}{k}} \cdot P_{y,n-1}, \; 0 \le j \le k$$