



## AML5103 | Applied Probability and Statistics | In-class Problem Set-1

- Consider a hash table with 5 buckets, where the probability of a string getting hashed to bucket  $i$  is given by  $p_i$  (where  $\sum_{i=1}^5 p_i = 1$ .) Now, 6 strings are hashed into the hash table. Determine the probability that each of the first 4 buckets has at least 1 string hashed to each of them. Explicitly expand your answer in terms of  $p_i$ , so that it does not include any summations.

### Solution:

- Let  $E$  be the event that each of buckets 1 to 4 has at least one string hashed to it.
- Let  $F_i$  be the event that bucket  $i$  has at least one string hashed to it.
- $E = F_1 \cap F_2 \cap F_3 \cap F_4$ . Interpret this in plain English.
- Are the events  $F_1, F_2, F_3, F_4$  intuitively independent or dependent? For example, if bucket-1 has all strings hashed to it, does it give any information about strings getting hashed to bucket-2?
- Interpret  $E^c$  in plain English.
- $P(E) = 1 - P(E^c) = 1 - P((F_1 \cap F_2 \cap F_3 \cap F_4)^c) = 1 - P(? \cup ? \cup ? \cup ?)$  by DeMorgan's law.
- The inclusion-exclusion principle for 4 events is:

$$\begin{aligned}
 P(A_1 \cup A_2 \cup A_3 \cup A_4) = & \\
 & P(A_1) + P(A_2) + P(A_3) + P(A_4) \\
 & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_1 \cap A_4) - P(A_2 \cap A_3) - P(A_2 \cap A_4) - P(A_3 \cap A_4) \\
 & + P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_4) + P(A_1 \cap A_3 \cap A_4) + P(A_2 \cap A_3 \cap A_4) \\
 & - P(A_1 \cap A_2 \cap A_3 \cap A_4).
 \end{aligned}$$

- Observe how the  $+/-$  signs alternate in relation to the number of events that are intersecting.
- Use the above result to compute  $P(F_1^c \cup F_2^c \cup F_3^c \cup F_4^c)$  and the desired probability  $P(E)$ .
- In doing so, observe, for example, that:

$$\begin{aligned}
 P(F_1^c) &= P(\text{no string hashed to bucket 1}) = (1-p_1)^6, \\
 P(F_1^c \cap F_2^c) &= P(\text{no string hashed to buckets 1 to 2}) = (1-(p_1+p_2))^6,
 \end{aligned}$$

and so on, where we have used that the event “no string hashed to bucket-1” is equivalent to “string-1 not hashed to bucket-1 AND string-2 not hashed to bucket-1 AND string-3 not hashed to bucket-1 AND string-4 not hashed to bucket-1 AND string-5 not hashed to bucket-1 AND string-6 not hashed to bucket-1.” Are the AND-connected events here independent? Recall that each string gets independently hashed to any one of the buckets. Also note that the event “no string hashed to buckets 1 to 2” is equivalent to “no string is hashed to bucket-1 and bucket-2” which is in turn equivalent to “string-1 not hashed to bucket-1 or bucket-2 AND string-2 not hashed to bucket-1 or bucket-2 AND string-3 not hashed to bucket-1 or bucket-2 AND string-4 not hashed to bucket-1 or bucket-2 AND string-5 not hashed to bucket-1 or bucket-2 AND string-6 not hashed to bucket-1 or bucket-2.”