11.2-1.

Suppose we use a hash function h to hash n distinct keys into an array T of length m. Assuming simple uniform hashing, what is the expected number of collisions? More precisely, what is the expected cardinality of $\{\{k,l\}: k \neq l \text{ and } h(k) = h(l)\}$?

Answer.

For each pair of distinct keys k and l, we define the indicator random variable $X_{kl} = I\{h(k) = h(l)\}$. Under the assumption of simple uniform hashing, we have $\Pr\{h(k) = h(l)\} = 1/m$, and so by Lemma 5.1, $E[X_{kl}] = 1/m$. Let Y denoted the total number of collisions, so that $Y = \sum_{k \neq l} X_{kl}$. The expected number of collisions is

$$
\begin{aligned}
E[Y] &= E\left[\sum_{k \neq l} X_{kl}\right] \\
&= \sum_{k \neq l} E[X_{kl}] \quad \text{(by linearity of expectation)} \\
&= \binom{n}{2}\frac{1}{m} \\
&= \frac{n(n-1)}{2} \cdot \frac{1}{m} \\
&= \frac{n(n-1)}{2m}
\end{aligned}
$$