1 Universal Hashing

**Problem 1.** For any choice of hash function, there exists a bad set of keys that all hash to the same slot.

The idea for solving this problem is to choose hash function at random independently from keys.

For keys \((x_1, x_2)\) and randomly picked \(a_1, a_2 \in \{0, 1, \ldots, n-1\}\), define \(h_{a_1,a_2}(x_1, x_2) = a_1x_1 + a_2x_2 \pmod{n}\) and there is a family of finite functions \(H = \{h_{a_1,a_2} | a_1, a_2 \in \{0, 1, \ldots, n-1\}\}\).

**Definition 2.** A family \(H : \text{keys} \to \{0, 1, \ldots, n-1\}\) of hash functions is 2-universal if for all \(x \neq y\)

\[
P(\forall h \in H \ h(x) = h(y)) = \frac{1}{n}.
\]

2 Balls and Bins

**Problem 3.** Throw \(m\) balls into \(n\) bins independently and uniformly randomly.

- How many empty bins?
- How full is the fullest bin?
- What is the distribution of occupancies?

**Theorem 4.** If events \(E_1\) and \(E_2\) are independent, then

\[
P(E_1 \cap E_2) = P(E_1) P(E_2).
\]

\[
P[\text{ball i lands in bin j}] = \frac{1}{n};
\]

\[
\text{bin j empty: } \{\text{ball 1 does not land in bin j}\} \cap \{\text{ball 2 does not land in bin j}\} \cap \cdots \cap \{\text{ball m does not land in bin j}\};
\]

\[
P(\text{bin j empty}) = \prod_{i=1}^{m} P[\text{ball i does not land in bin j}] = (1 - \frac{1}{n})^m;
\]

If \(m = n\), \(P(\text{bin j empty}) = (1 - \frac{1}{n})^n \to \frac{1}{e} \approx 0.37\).
**Definition 5.** A random variable is defined on a set of possible outcomes (the sample space $\Omega$) and a probability distribution that associate each outcome with a probability.

**Example 6.** Define random variable $X$ on the result of coin toss

$$X = \begin{cases} 0 & \text{head} \\ 1 & \text{tail} \end{cases}$$

and associate probability $P(X = 0) = \frac{1}{2}$, $P(X = 1) = \frac{1}{2}$.

Another interesting example is related to the probability of two mutually exclusive events.

**Example 7.** Assume the random variable $X \in \{0, 1, \cdots, 9\}$ has distribution $P(X = i) = \frac{1}{10}$. Then

$$P(x^2 - 3x + 2 = 0) = P(x = 1) + P(x = 2) = \frac{1}{10} + \frac{1}{10} = \frac{1}{5}.$$

**Definition 8.** The expectation of a random variable is the weighted average of all possible values that this random variable can take on

$$E[X] = \sum_{x \in \Omega} x \cdot P(X = x).$$

**Example 9.** Given random variable and corresponding probability distribution

$$P(X = 0) = \frac{1}{2}, \ P(X = 1) = \frac{1}{4}, \ P(X = 2) = \frac{1}{4}$$

then the expectation is

$$E[x] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} = \frac{3}{4}.$$

**Theorem 10.** The linearity of expectation property guarantees that for any random variables $X$ and $Y$, we have

$$E[X + Y] = E[X] + E[Y].$$

Proof: $E[X + Y] = \sum_{z} x \cdot P(X + Y = z) = \sum_{x,y} (x + y) \cdot P(X = x \cap Y = y) = \sum_{x} \sum_{y} x P(X = x \cap Y = y) + \sum_{y} y P(X = x \cap Y = y) = \sum_{x} x P(X = x) + \sum_{y} y P(Y = y) = E[x] + E[y]$ □

Define

$$X_i = \begin{cases} 1 & \text{if bin } i \text{ is empty} \\ 0 & \text{else} \end{cases}$$

then, we have $X = X_1 + \cdots + X_n$ is the number of empty bins

$$E[\# \text{ empty bins}] = E[X] = \sum_{i=1}^{n} E[X_i] = \frac{n}{e}.$$