Reservoir Sampling

Justin Alvin & Prakhar Sharma

Base Case: $n = s$
We fill the reservoir with all s samples.

$$P(\text{Subset } s \text{ in reservoir}) = \frac{1}{\binom{n}{s}} = \frac{1}{\binom{s}{s}} = 1$$

Induction Hypothesis: This holds for all $n = i$

$$P(\text{Subset } s \text{ in reservoir}) = \frac{1}{\binom{i}{s}}$$

For $n = i$:
Case I: Subset s is already in reservoir and element i is discarded.

$$P(i^{th}\text{element selected}) = \frac{s}{i}$$

$$P(i^{th}\text{element discarded}) = \frac{i-s}{i} = 1 - \frac{s}{i}$$

$$P(\text{Subset } s \text{ survives}) = P(i^{th}\text{element discarded}) \times P(i^{th}\text{element selected})$$

$$= (1 - \frac{s}{i}) \cdot \frac{1}{\binom{i-1}{s}} = \frac{s!(i-1-s)!}{(i-1)!} \cdot \frac{i-s}{i} = \frac{s!(i-s)!}{i!} = \frac{1}{\binom{i}{s}}$$

Case II: Element i is selected and is in subset s. All elements of s but i are already in reservoir and element j is in the reservoir but not part of subset s.

$$P(\text{All elements but } i \text{ in reservoir}) = (i-s) \cdot \frac{1}{\binom{i-1}{s}}$$

$$P(\text{Select element } i \text{ and discard element } j) = \frac{s}{i} \cdot \frac{1}{s}$$

$$P(\text{Subset } s \text{ in reservoir}) = P(\text{All elements but } i \text{ in res.}) \times P(\text{Select element } i \text{ and discard } j)$$

$$= (i-s) \cdot \frac{1}{\binom{i-1}{s}} \cdot \frac{s}{i} \cdot \frac{1}{s} = \frac{s!(i-1-s)!}{(i-1)!} \cdot \frac{i-s}{i} = \frac{s!(i-s)!}{i!} = \frac{1}{\binom{i}{s}}$$