A variant of coupon collection problem*

Parimal Parag †

Nov 18, 2023

1 System Model

Consider a variant of the coupon collection problem, where there are R replicas each for balls of V different colors. There are B bins, and each can store K balls, such that $VR = BK$. In each round, one can sample one of the B balls uniformly at random, independent of past rounds. A person collects a ball from the sampled bin if it has a ball with a color distinct from all the ones it has already collected. If the bin doesn't have any such ball to offer, the round is wasted. We stop after all distinct V colors of balls are collected.

We denote by $\Phi : [V] \to \mathcal{P}([B])$ the storage mapping, such that Φ_v is the set of bins having balls of color v. Accordingly, $S : [B] \to \mathcal{P}([V])$ is the color mapping, such that S_b is the set of colors of balls with bin $b \in [B]$. After collection of ℓ balls of distinct colors, we denote the I_ℓ to be the random set of ℓ collected colors. After the collection of ℓ distinct colored balls, the choice of color at bin b is denoted by $\Pi_b : I_\ell \to S_b \setminus I_\ell$. Let T_ℓ be the number of rounds to collect the ℓth distinct colored ball after the collection of the $\ell - 1$ distinct colored balls.

Problem 1. Find the optimal storage mapping $\Phi : [V] \to \mathcal{P}([B])$ and collection mapping $\Pi : [B] \to \mathcal{P}([V])^{\mathcal{P}([V])}$ such that the mean time to collect all distinct colored balls averaged over all colors $\frac{1}{V}\mathbb{E}\sum_{\ell=1}^{V} T_\ell$ is minimized.

*This presentation is based on [1].

†Rooji Jinan was with the Department of Cyber-Physical Systems, Indian Institute of Science, Bangalore, Karnataka 560012, India. Ajay Badita and Parimal Parag were with the Department of Electrical and Communications Engineering, Indian Institute of Science, Bangalore, Karnataka 560012, India. Pradeep Kiran Sarvepalli is with the Department of Electrical Engineering, Indian Institute of Technology Madras, TN 600036, India. Emails: {roojijinan, ajaybadita, parimal}@iisc.ac.in, pradeep@ee.iitm.ac.in.

2 Analysis

We denote the set of bins with uncollected colored balls after ℓ distinct colored ball collections by $U(I_\ell) \triangleq \cup_{v \notin I_\ell} \Phi_v$ and referred to as *useful bins*. The number of useful bins is denoted by $N_\ell \triangleq |U(I_\ell)|$ after collecting ℓ balls of distinct colors. We observe that each bin can be chosen with a uniform probability on each round. However, after collection of I_ℓ colors, sampling a bin in $[B] \setminus U(I_\ell)$ leads to wasting of a round. It follows that the number of rounds $T_{\ell+1}$ to collect $(\ell+1)$th ball of distinct color is geometric with mean $\frac{B}{N_\ell}$. It follows that the mean time to collect balls of all colors is

$$\frac{1}{V}\mathbb{E}\sum_{\ell=1}^{V} T_\ell = \frac{1}{V}\mathbb{E}\sum_{\ell=1}^{V}\mathbb{E}[T_\ell \mid I_{\ell-1}] = \frac{B}{V}\mathbb{E}\sum_{\ell=0}^{V-1}\frac{1}{N_\ell}. \tag{1}$$

We observe that the expectation is over the randomness in choice of subset of colors I_ℓ that depends on the storage mapping Φ and the color choice function Π.

Example 1 (Identical replicas at each bin)**.** Consider the case when $V = B$ and $R = K$, and each bin has all R balls of the same color. In this case, we can assume without any loss of generality that $\Phi_v = \{v\}$ for all $v \in [V]$ and $S_b = \{b\}$ for all $b \in [B]$. Further, we have $U(I_\ell) = [V] \setminus I_\ell^c$. It follows that the mean collection time is

$$\sum_{\ell=0}^{V-1}\frac{1}{V-\ell}.$$

We have $I_\ell \in \binom{[V]}{\ell} \triangleq \{S \subseteq [V] : |S| = \ell\}$ with uniform probability, and hence we note that the mean collection time is independent of the color choice mapping Π. In addition, we observe that N_ℓ reduces by a unit amount after every collection, and hence this is not a good storage scheme.

In general, we would focus on storage maps that allocate distinct colors at each bin. In this case, we have

$$\Phi : [V] \to \binom{[B]}{R}, \qquad S : [B] \to \binom{[V]}{K}, \qquad S_b \triangleq \cup_{v\in[V]}(\Phi_v \cap \{b\}).$$

Lemma 1. *The mean collection time for V balls of distinct colors is lower bounded by*

$$\frac{1}{V}\mathbb{E}\sum_{\ell=1}^{V} T_\ell \geqslant \frac{BV}{\mathbb{E}\sum_{\ell=0}^{V-1} N_\ell}.$$

Proof. From the AM-HM inequality in (1), we observe that

$$\frac{1}{V}\mathbb{E}\sum_{\ell=1}^{V} T_\ell = \frac{B}{V}\mathbb{E}\sum_{\ell=0}^{V-1}\frac{1}{N_\ell} \geqslant B\mathbb{E}\frac{V}{\sum_{\ell=0}^{V-1} N_\ell}.$$

From the convexity of the map $x \mapsto 1/x$, and the application of Jensen's inequality to this convex map, we obtain

$$\mathbb{E}\frac{V}{\sum_{\ell=0}^{V-1} N_\ell} \geqslant \frac{V}{\mathbb{E}\sum_{\ell=0}^{V-1} N_\ell}.$$

□

Instead of minimizing the mean completion time, we focus on minimizing its lower bound, which is equivalent to maximizing the mean number of useful bins. We ask the following simpler problem.

Problem 2. Find the optimal storage mapping $\Phi : [V] \to \binom{[B]}{R}$ and collection mapping $\Pi : [B] \to \binom{[V]}{K}^{\binom{[V]}{K}}$ such that the mean number of distinct colors averaged over all colors $\frac{1}{V}\mathbb{E}\sum_{\ell=0}^{V-1} N_\ell$ is maximized.

Example 2 (Cyclic shifted storage)**.** Consider the case when $V = B$ and $R = K$ and $\Phi_v = \{v, v+1, \dots, v+R-1\}$ for all $v \in [V]$ where the addition is modulo V. In addition, we observe that N_ℓ reduces by unit amount after every collection, and hence this is also not a good storage scheme.

References

[1] R. Jinan, A. Badita, P. K. Sarvepalli, and P. Parag, "Latency optimal storage and scheduling of replicated fragments for memory constrained servers," *IEEE Trans. Inf. Theory*, vol. 68, no. 6, pp. 4135–4155, Jun. 2022.