Chapter 9

Coupon's Collector Problems II

By Sariel Har-Peled, April 26, 2022[①]

> There is not much talking now. A silence falls upon them all. This is no time to talk of hedges and fields, or the beauties of any country. Sadness and fear and hate, how they well up in the heart and mind, whenever one opens the pages of these messengers of doom. Cry for the broken tribe, for the law and custom that is gone. Aye, and cry aloud for the man who is dead, for the woman and children bereaved. Cry, the beloved country, these things are not yet at an end. The sun pours down on the earth, on the lovely land that man cannot enjoy. He knows only the fear of his heart.
>
> Alan Paton, Cry, the beloved country

9.1. The Coupon Collector's Problem Revisited

9.1.1. Some technical lemmas

Unfortunately, in Randomized Algorithms, many of the calculations are awful[②]. As such, one has to be dexterous in approximating such calculations. We present quickly a few of these estimates.

Lemma 9.1.1. *For $x \geq 0$, we have $1 - x \leq \exp(-x)$ and $1 + x \leq e^x$. Namely, for all x, we have $1 + x \leq e^x$.*

Proof: For $x = 0$ we have equality. Next, computing the derivative on both sides, we have that we need to prove that $-1 \leq -\exp(-x) \iff 1 \geq \exp(-x) \iff e^x \geq 1$, which clearly holds for $x \geq 0$.

A similar argument works for the second inequality. ■

Lemma 9.1.2. *For any $y \geq 1$, and $|x| \leq 1$, we have $\left(1 - x^2\right)^y \geq 1 - yx^2$.*

Proof: Observe that the inequality holds with equality for $x = 0$. So compute the derivative of x of both sides of the inequality. We need to prove that

$$y(-2x)\left(1 - x^2\right)^{y-1} \geq -2yx \iff \left(1 - x^2\right)^{y-1} \leq 1,$$

which holds since $1 - x^2 \leq 1$, and $y - 1 \geq 0$. ■

Lemma 9.1.3. *For any $y \geq 1$, and $|x| \leq 1$, we have $\left(1 - x^2 y\right)e^{xy} \leq (1 + x)^y \leq e^{xy}$.*

Proof: The right side of the inequality is standard by now. As for the left side. Observe that

$$(1 - x^2)e^x \leq 1 + x,$$

since dividing both sides by $(1+x)e^x$, we get $1 - x \leq e^{-x}$, which we know holds for any x. By Lemma 9.1.2, we have

$$\left(1 - x^2 y\right)e^{xy} \leq \left(1 - x^2\right)^y e^{xy} = \left(\left(1 - x^2\right)e^x\right)^y \leq \left(1 + x\right)^y \leq e^{xy}.$$ ■

[②]"In space travel," repeated Slartibartfast, "all the numbers are awful." – Life, the Universe, and Everything Else, Douglas Adams.

9.1.2. Back to the coupon collector's problem

There are n types of coupons, and at each trial one coupon is picked in random. How many trials one has to perform before picking all coupons? Let m be the number of trials performed. We would like to bound the probability that m exceeds a certain number, and we still did not pick all coupons.

In the previous lecture, we showed that

$$\mathbb{P}\left[\#\text{ of trials} \geq n\log n + n + t\cdot n\frac{\pi}{\sqrt{6}}\right] \leq \frac{1}{t^2},$$

for any t.

A stronger bound, follows from the following observation. Let Z_i^r denote the event that the ith coupon was not picked in the first r trials. Clearly,

$$\mathbb{P}\left[Z_i^r\right] = \left(1 - \frac{1}{n}\right)^r \leq \exp\left(-\frac{r}{n}\right).$$

Thus, for $r = \beta n \log n$, we have $\mathbb{P}\left[Z_i^r\right] \leq \exp\left(-\frac{\beta n \log n}{n}\right) = n^{-\beta}$. Thus,

$$\mathbb{P}\left[X > \beta n \log n\right] \leq \mathbb{P}\left[\bigcup_i Z_i^{\beta n \log n}\right] \leq n \cdot \mathbb{P}\left[Z_1\right] \leq n^{-\beta+1}.$$

Lemma 9.1.4. *Let the random variable X denote the number of trials for collecting each of the n types of coupons. Then, we have $\mathbb{P}\left[X > n\ln n + cn\right] \leq e^{-c}$.*

Proof: The probability we fail to pick the first type of coupon is $\alpha = (1 - 1/n)^m \leq \exp\left(-\frac{n\ln n + cn}{n}\right) = \exp(-c)/n$. As such, using the union bound, the probability we fail to pick all n types of coupons is bounded by $n\alpha = \exp(-c)$, as claimed. ■

In the following, we show a slightly stronger bound on the probability, which is $1 - \exp(-e^{-c})$. To see that it is indeed stronger, observe that $e^{-c} \geq 1 - \exp(-e^{-c})$.

9.1.3. An asymptotically tight bound

Lemma 9.1.5. *Let $c > 0$ be a constant, $m = n\ln n + cn$ for a positive integer n. Then for any constant k, we have $\lim_{n\to\infty} \binom{n}{k}\left(1 - \frac{k}{n}\right)^m = \frac{\exp(-ck)}{k!}$.*

Proof: By Lemma 9.1.3, we have

$$\left(1 - \frac{k^2 m}{n^2}\right)\exp\left(-\frac{km}{n}\right) \leq \left(1 - \frac{k}{n}\right)^m \leq \exp\left(-\frac{km}{n}\right).$$

Observe also that $\lim_{n\to\infty}\left(1 - \frac{k^2 m}{n^2}\right) = 1$, and $\exp\left(-\frac{km}{n}\right) = n^{-k}\exp(-ck)$. Also,

$$\lim_{n\to\infty} \binom{n}{k}\frac{k!}{n^k} = \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1.$$

Thus, $\lim_{n\to\infty} \binom{n}{k}\left(1 - \frac{k}{n}\right)^m = \lim_{n\to\infty} \frac{n^k}{k!}\exp\left(-\frac{km}{n}\right) = \lim_{n\to\infty} \frac{n^k}{k!} n^{-k}\exp(-ck) = \frac{\exp(-ck)}{k!}$. ■

Theorem 9.1.6. *Let the random variable X denote the number of trials for collecting each of the n types of coupons. Then, for any constant $c \in \mathbb{R}$, and $m = n \ln n + cn$, we have $\lim_{n\to\infty} \mathbb{P}\left[X > m\right] = 1 - \exp\left(-e^{-c}\right)$.*

Before dwelling into the proof, observe that $1 - \exp(-e^{-c}) \approx 1 - (1 - e^{-c}) = e^{-c}$. Namely, in the limit, the upper bound of Lemma 9.1.4 is tight.

Proof: We have $\mathbb{P}\left[X > m\right] = \mathbb{P}\left[\cup_i Z_i^m\right]$. By inclusion-exclusion, we have

$$\mathbb{P}\left[\bigcup_i Z_i^m\right] = \sum_{i=1}^{n} (-1)^{i+1} P_i^n,$$

where $P_j^n = \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} \mathbb{P}\left[\cap_{v=1}^{j} Z_{i_v}^m\right]$. Let $S_k^n = \sum_{i=1}^{k} (-1)^{i+1} P_i^n$. We know that $S_{2k}^n \le \mathbb{P}\left[\bigcup_i Z_i^m\right] \le S_{2k+1}^n$.

By symmetry,

$$P_k^n = \binom{n}{k} \mathbb{P}\left[\bigcap_{v=1}^{k} Z_v^m\right] = \binom{n}{k}\left(1 - \frac{k}{n}\right)^m,$$

Thus, $P_k = \lim_{n\to\infty} P_k^n = \exp(-ck)/k!$, by Lemma 9.1.5. Thus, we have

$$S_k = \sum_{j=1}^{k} (-1)^{j+1} P_j = \sum_{j=1}^{k} (-1)^{j+1} \cdot \frac{\exp(-cj)}{j!}.$$

Observe that $\lim_{k\to\infty} S_k = 1 - \exp(-e^{-c})$ by the Taylor expansion of $\exp(x)$ (for $x = -e^{-c}$). Indeed,

$$\exp(x) = \sum_{j=0}^{\infty} \frac{x^j}{j!} = \sum_{j=0}^{\infty} \frac{(-e^{-c})^j}{j!} = 1 + \sum_{j=1}^{\infty} \frac{(-1)^j \exp(-cj)}{j!}.$$

Clearly, $\lim_{n\to\infty} S_k^n = S_k$ and $\lim_{k\to\infty} S_k = 1 - \exp(-e^{-c})$. Thus, (using fluffy math), we have

$$\lim_{n\to\infty} \mathbb{P}\left[X > m\right] = \lim_{n\to\infty} \mathbb{P}\left[\cup_{i=1}^{n} Z_i^m\right] = \lim_{n\to\infty} \lim_{k\to\infty} S_k^n = \lim_{k\to\infty} S_k = 1 - \exp(-e^{-c}).$$ ■

9.2. Bibliographical notes

Are presentation follows, as usual, Motwani and Raghavan [MR95].

References

[MR95] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge, UK: Cambridge University Press, 1995.