CS 574: Randomized Algorithms

Lecture 5. Coupon Collector Problems

September 8, 2015

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.
- We now show a weak bound using Chebyshev, stronger bounds later.

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.
- We now show a weak bound using Chebyshev, stronger bounds later.
- Show that for an r.v. Y with $geom(p)$ distribution, $E(Y) = \frac{1}{p}$ and $Var(Y) = \frac{(1-p)}{p^2}$.

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.
- We now show a weak bound using Chebyshev, stronger bounds later.
- Show that for an r.v. Y with $geom(p)$ distribution, $E(Y) = \frac{1}{p}$ and $Var(Y) = \frac{(1-p)}{p^2}$.
- For any t that

$$Pr[\#boxes \geq n \log n + n + t \cdot n \frac{\pi}{\sqrt{6}}] \leq \frac{1}{t^2}$$

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.
- We now show a weak bound using Chebyshev, stronger bounds later.
- Show that for an r.v. Y with $geom(p)$ distribution, $E(Y) = \frac{1}{p}$ and $Var(Y) = \frac{(1-p)}{p^2}$.
- For any t that

$$Pr[\#boxes \geq n \log n + n + t \cdot n \frac{\pi}{\sqrt{6}}] \leq \frac{1}{t^2}$$

- Can you cast it in Balls-in-Bins framework?

Coupon Collector

- n types of coupons in cereal boxes,each time you purchase a cereal box, one coupon is picked at random. How many boxes one has to buy before picking all coupons?
- m is the number of cereal boxes. We want to bound the probability that m exceeds a certain number and we still did not pick all coupons.
- We now show a weak bound using Chebyshev, stronger bounds later.
- Show that for an r.v. Y with $geom(p)$ distribution, $E(Y) = \frac{1}{p}$ and $Var(Y) = \frac{(1-p)}{p^2}$.
- For any t that

$$Pr[\#boxes \geq n \log n + n + t \cdot n \frac{\pi}{\sqrt{6}}] \leq \frac{1}{t^2}$$

- Can you cast it in Balls-in-Bins framework?

Coupon Collector, Revisited

- What is the probability that the i-th coupon was not picked the first r trials? (event ${E_i}^r$)

Coupon Collector, Revisited

- What is the probability that the i-th coupon was not picked the first r trials? (event ${E_i}^r$)
- Stronger bound than before: $Pr[X > \beta n \log n] \leq n^{-\beta+1}$.

Coupon Collector, Revisited

- What is the probability that the i-th coupon was not picked the first r trials? (event $E_i{}^r$)
- Stronger bound than before: $Pr[X > \beta n \log n] \leq n^{-\beta+1}$.We can do even better concentration for the probability that X deviates from its expectation nH_n by cn.

Coupon Collector, Revisited

- What is the probability that the i-th coupon was not picked the first r trials? (event $E_i{}^r$)
- Stronger bound than before: $Pr[X > \beta n \log n] \leq n^{-\beta+1}$.We can do even better concentration for the probability that X deviates from its expectation nH_n by cn.

Theorem

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

$$\lim_{n \to \infty} Pr[X > m] = 1 - exp(-e^{-c})$$

Coupon Collector, Revisited

- What is the probability that the i-th coupon was not picked the first r trials? (event $E_i{}^r$)
- Stronger bound than before: $Pr[X > \beta n \log n] \leq n^{-\beta+1}$.We can do even better concentration for the probability that X deviates from its expectation nH_n by cn.

Theorem

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

$$\lim_{n \to \infty} Pr[X > m] = 1 - exp(-e^{-c})$$

How many cereal boxes and the Poisson Heuristic

- Observe that as c goes from large positive to large negative value, the probability goes from almost 1 to almost 0. So if you have collected almost $n \log n$ cereal boxes, don't give up!

How many cereal boxes and the Poisson Heuristic

- Observe that as c goes from large positive to large negative value, the probability goes from almost 1 to almost 0. So if you have collected almost $n \log n$ cereal boxes, don't give up!
- We will prove an approximate version of that, using Poisson approximation.

How many cereal boxes and the Poisson Heuristic

- Observe that as c goes from large positive to large negative value, the probability goes from almost 1 to almost 0. So if you have collected almost $n \log n$ cereal boxes, don't give up!
- We will prove an approximate version of that, using Poisson approximation.

Poisson Heuristic

- ${N_i}^r$ is number of times coupon i i selected the first r trials. Follows Binomial$(r, p = 1/n)$.

Poisson Heuristic

- ${N_i}^r$ is number of times coupon i i selected the first r trials. Follows Binomial$(r, p = 1/n)$.
- $Pr[{N_i}^r = x] = \binom{r}{x} p^x (1-p)^{r-x}$.

Poisson Heuristic

- ${N_i}^r$ is number of times coupon i i selected the first r trials. Follows Binomial$(r, p = 1/n)$.
- $Pr[{N_i}^r = x] = \binom{r}{x} p^x (1-p)^{r-x}$.
- An rv. Y follows Poison with parameter λ if $Pr[Y = x] = \frac{\lambda^y e^{-\lambda}}{y!}$.

Poisson Heuristic

- $N_i{}^r$ is number of times coupon i i selected the first r trials. Follows Binomial$(r, p = 1/n)$.
- $Pr[N_i{}^r = x] = \binom{r}{x} p^x (1-p)^{r-x}$.
- An rv. Y follows Poison with parameter λ if $Pr[Y = x] = \frac{\lambda^y e^{-\lambda}}{y!}$.
- $Poisson(\lambda = rp) \approx Binomial(r, p)$.

Poisson Heuristic

- ${N_i}^r$ is number of times coupon i i selected the first r trials. Follows Binomial$(r, p = 1/n)$.
- $Pr[{N_i}^r = x] = \binom{r}{x} p^x (1-p)^{r-x}$.
- An rv. Y follows Poison with parameter λ if $Pr[Y = x] = \frac{\lambda^y e^{-\lambda}}{y!}$.
- $Poisson(\lambda = rp) \approx Binomial(r, p)$.
- Now the events ${E_i}^r$ almost independent.

Class Assignment: Overview of Techniques

Unbalancing lights: Consider a square $n \times n$ array of lights (see Figure on board). There is one switch corresponding to each row and each column (i.e., $2n$ switches). Throwing a switch changes the state of all the lights in the corresponding row or column. We now consider the problem of setting the switches so as to maximize the number of lights that are ON, starting from an arbitrary configuration of switches. You need to show the following claim:

Claim

For any initial configuration of the lights, there exists a setting of the switches for which the number of lights that are on is asymptotically

$$\frac{n^2}{2} + \sqrt{\frac{1}{2\pi}}n^{3/2}$$

Class Assignment: Overview of Techniques

Unbalancing lights: Consider a square $n \times n$ array of lights (see Figure on board). There is one switch corresponding to each row and each column (i.e., $2n$ switches). Throwing a switch changes the state of all the lights in the corresponding row or column. We now consider the problem of setting the switches so as to maximize the number of lights that are ON, starting from an arbitrary configuration of switches. You need to show the following claim:

Claim

For any initial configuration of the lights, there exists a setting of the switches for which the number of lights that are on is asymptotically

$$\frac{n^2}{2} + \sqrt{\frac{1}{2\pi}} n^{3/2}$$

As an intermediate step, show that in expectation we achieve about $\frac{n^2}{2} + O(n)$ lights on.