CS5314
Randomized Algorithms

Lecture 14: Balls, Bins, Random Graphs
(Poisson Approximation)

Objectives

- Poisson Approximation for Balls-and-Bins :
  to approximate # balls in each bin as independent Possion RV with $\mu = m/n$

- Revisit Coupon Collector

Poisson Approximation

- Suppose we throw m balls into n bins independently and uniformly at random

- From previous lecture, we observe that:

  # balls in a particular bin
  $\sim$ Poisson RV with $\mu = m/n$

- How about distribution of balls in all n bins ?

Poisson Approximation

Question: Will distribution of n balls be the same as n independent Poisson RVs with mean m/n?

Ans. No !

For instance, total # of balls is always exactly m, but sum of n independent Poisson RVs can be any value

The difference is because of dependency !

Poisson Approximation

- Though

  "n independent Poisson RVs"

  do not have the same distribution as

  "m balls into n bins"

  we can show that they are related, so that we can use the "Poisson Case" to approximate the "Exact Case"

Hopefully, the approximation will be useful …

Poisson Approximation

Formally, we define

$$X_1^{(m)}, X_2^{(m)}, \ldots, X_n^{(m)}$$

where $X_j^{(m)}$ = # balls in Bin j (in Exact Case)

$$Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_n^{(m)}$$

which are n independent Poisson RVs with parameter m/n (in Poisson Case)

When Two Distributions Meet

Theorem: Suppose $\sum_{j=1 \text{ to } n} Y_j^{(m)} = k$. Under this condition, the distribution of

$$(Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_n^{(m)})$$

is exactly the same as the distribution of

$$(X_1^{(k)}, X_2^{(k)}, \ldots, X_n^{(k)})$$

regardless of the value of m or k

Throwing k balls in total

How to prove?

Proof

- Let $k_1, k_2, \ldots, k_n$ be non-negative integers whose sum is k

- When throwing k balls into n bins,

$$\Pr\left((X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n) \right)$$

$$= \frac{k\,!}{k_1!\, k_2! \cdots k_n!\, n^k}$$

Proof

Next,

$$\Pr\Big((Y_1^{(m)}, \ldots, Y_n^{(m)}) = (k_1, \ldots, k_n) \mid \textstyle\sum_j Y_j^{(m)} = k\Big)$$

$$= \frac{\Pr((Y_1^{(m)} = k_1) \cap \cdots \cap (Y_n^{(m)} = k_n))}{\Pr(\sum_j Y_j^{(m)} = k)} \quad \ldots \text{(why?)}$$

Question: What is this probability??

Proof

First,

$$\Pr(Y_j^{(m)} = k_j) = e^{-m/n}(m/n)^{k_j} / k_j!$$

Since $Y_1^{(m)}, \ldots, Y_n^{(m)}$ are independent, so

$$\Pr((Y_1^{(m)} = k_1) \cap \cdots \cap (Y_n^{(m)} = k_n))$$

$$= \prod_j e^{-m/n}(m/n)^{k_j} / k_j!$$

$$= \frac{e^{-m}\, m^k}{k_1!\, k_2! \cdots k_n!\, n^k}$$

Proof

On the other hand,

$$Pr(\Sigma_j Y_j^{(m)} = k) = e^{-m} m^k / k! \qquad \text{... [why??]}$$

So combining the previous results,

$$Pr((Y_1^{(m)}, \ldots, Y_n^{(m)}) = (k_1, \ldots, k_n) \mid \Sigma_j Y_j^{(m)} = k)$$

$$= Pr((X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n))$$

➔ this completes the proof

A Stronger Result

- With the previous result between exact case and Poisson case, we can show a stronger result …

- Before we proceed, let us obtain a useful upper bound for n !

Upper Bound for n!

Lemma: $n! \leq e n^{1/2} (n/e)^n$

Proof: Since ln x is a concave function,

$$\int_{j-1}^{j} \ln x \, dx \geq (\ln(j-1) + \ln j) / 2 \quad \text{... (why?)}$$

➔ $$\int_{1}^{n} \ln x \, dx \geq \ln(n!) - (\ln n)/2 \quad \text{... (why?)}$$

➔ $n \ln n - n + 1 \geq \ln(n!) - (\ln n)/2$

➔ Lemma follows by exponentiation

Expectation of Loads

- We now show a relationship between the expectation of any non-negative function of the loads in the two cases :

Theorem:

Let $f(x_1, \ldots, x_n)$ be a non-negative function.

Then,

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \leq e\sqrt{m}\, E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$

How to prove?

Proof

$$E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$

$$= \sum_k E[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) \mid \Sigma_j \; Y_j^{(m)}=k] \; Pr(\Sigma_j \; Y_j^{(m)}=k)$$

$$\geq E[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) \mid \Sigma_j Y_j^{(m)}=m] \; Pr(\Sigma_j Y_j^{(m)}=m)$$

$$= E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \; Pr(\Sigma_j Y_j^{(m)}=m)$$ … (why?)

Proof

Next, using upper bound of m! ,

$$\Pr(\Sigma_j Y_j^{(m)} = m) = e^{-m}\, m^m / m! \qquad \text{... (why?)}$$
$$\geq 1 / (e m^{1/2})$$

Thus,

$$E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$
$$\geq E[f(X_1^{(m)}, \ldots, X_n^{(m)})] / (e m^{1/2})$$

➔ This completes the proof

Remark

- The previous theorem holds for any non-negative function f
- E.g., if f = MAX, then we can relate the expected maximum load in the two cases
- E.g., if f = an indicator for an event Z, then the theorem gives the relationship of Pr(Z occurs) in the two cases

This latter gives the following corollary:

Bounding Exact Case

Corollary: Referring to the scenario of throwing m balls into n bins.

Any event Z that takes place with probability p in the Poisson case implies: Z takes place with probability at most $em^{1/2}p$ in the exact case

How to prove?

Bounding Exact Case

Proof: Let f be the indicator for event Z

Then,

Pr(Z occurs in exact case)

$= E[f(X_1^{(m)}, \ldots, X_n^{(m)})]$

$\leq em^{1/2} E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$

$= em^{1/2}$ Pr(Z occurs in Poisson case)

$= em^{1/2}p$

An Even Stronger Result

If we know more about f, we can obtain an even stronger bound:

Theorem: Let $f(x_1, \ldots, x_n)$ be a non-negative function such that $E[f(X_1^{(m)}, \ldots, X_n^{(m)})]$ is monotonically increasing in m.

Then,

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \leq 2\,E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$

How to prove? (Ex. 5.13, 5.14)

Bounding Exact Case (2)

Corollary:

Let Z be an event whose probability is monotonically increasing in # balls.

If Z has probability p in the Poisson case,
→ Z has probability at most 2p in the exact case

Maximum Load (Revisited)

- Some time ago, we have shown that for sufficiently large n, if we throw n balls into n bins, then w.h.p. :

  Maximum load $\leq$ 3 ln n / ln ln n

- The proof is simply based on counting and union bound

- Let's see how the latest result can help in giving a lower bound…

Maximum Load (Revisited)

Lemma:

Suppose n balls are thrown to n bins, independently and uniformly at random.

Then w.h.p. (at least $1-1/n$) :

$$\text{Maximum load} \geq \ln n / \ln \ln n$$

How to prove?

Let's bound the probability for the Poisson case, and then…

Proof

Let $M = \ln n / \ln \ln n$

In the Poisson case,

$\Pr(\text{\# of balls in Bin } 1 \geq M)$

$\geq \Pr(\text{\# of balls in Bin } 1 = M)$

$= e^{-1}(1)^M / M! = 1/(eM!)$

➔ In the Poisson case,

$\Pr(\text{Max-Load} < M) \leq (1 - 1/(eM!))^n$

$\leq \exp\{ -n/(eM!) \}$

Proof

Next, we simplify the bound by showing:

$$- n / (eM!) \leq - c \ln n \qquad \text{for some } c$$

Recall that

$$M! \leq eM^{1/2} (M/e)^M$$

$$\leq M (M/e)^M \qquad \text{[for large n]}$$

$$\Rightarrow \ln M! \leq \ln M + M \ln M - M$$

$$\leq \ln \ln n + \ln n - M$$

$$\leq \ln n - \ln \ln n - \ln (2e) \qquad \text{[for large n]}$$

Proof

Thus,

$$M! \leq n / (2e \ln n) \qquad \text{[for large n]}$$

➔ $\exp\{ - n / (eM!) \} \leq \exp\{ -2\ln n \} = 1/n^2$

So, in the Poisson case

$$\Pr(\text{Max-Load} < M) \leq 1/n^2$$

➔ In the Exact case

$$\Pr(\text{Max-Load} < M) \leq e n^{1/2}(1/n^2) \leq 1/n$$

Coupon Collector (Revisited)

- Previously we have shown that if we want to collect a set of n coupons, the expected number of coupons we buy is

$$n\,H(n) \approx n \ln n$$

- Suppose we have bought $n \ln n + cn$ coupons already. What is the probability that we have obtained a full collection ?

Coupon Collector (Revisited)

- After buying $n \ln n + cn$ coupons:

  Pr(not having i^{th} coupon)

  $= (1 - 1/n)^{n \ln n + cn}$

  $\leq e^{-(1/n)(n \ln n + cn)} = e^{-c} / n$

- After buying $n \ln n + cn$ coupons:

  Pr(not having a full collection) $\leq e^{-c}$

➔ Pr(having a full collection) $\geq 1 - e^{-c}$

Coupon Collector (Revisited)

- Recently, we have seen that Chernoff bound usually gives a much tighter result

Question:

Can we apply Chernoff bound to get an even better result ?

Coupon Collector (Revisited)

Theorem: Let X be the number of coupons we buy before getting one card of each n types of coupons. Then, for any c,

$$\lim_{n\to\infty} \Pr(X > n \ln n + cn) = 1 - e^{-e^{-c}}$$

Remark: When $c = -4$, $1 - e^{-e^{-c}} \approx 1$

When $c = 4$, $1 - e^{-e^{-c}} \approx 0.02$

➔ For large n, #coupons is between $n \ln n \pm 4n$ is ~ 98% !!!

➔ This is an example of sharp threshold, where the random variable's distribution is concentrated around its mean

Proof

- We can consider the coupon collector's problem as a balls-and-bins problem (What are the balls? How many bins?)
- We shall use Poisson approximation so that intermediate steps will be easier
- Suppose # balls in each bin is a Poisson RV with mean $\ln n + c$, so that the expected total # balls is $m = n \ln n + cn$

Proof

Then, in the Poisson case,

$Pr(\text{Bin 1 is empty}) = e^{-(\ln n + c)} = e^{-c}/n$

Let NE be the event that no bin is empty in Poisson case

So, $Pr(NE) = (1 - e^{-c}/n)^n$

$= e^{-e^{-c}}$ … [when $n \to \infty$]

Two Facts

Let Y be # balls thrown in the Poisson case

Let $r = \sqrt{2m \ln m}$

We claim that as $n \to \infty$,

1. $\Pr(|Y-m| > r) = 0$ (i.e., Y is very close to mean)
2. $\Pr(NE \mid |Y-m| \leq r) = \Pr(NE \mid Y=m)$

   In case Y is very close to mean, we can just assume Y = m when computing Pr(NE)

Suppose our claim is true …

Consequence of Two Facts

As $n \to \infty$,

$$e^{-e^{-c}} = \Pr(NE)$$

$$= \Pr(NE \mid |Y-m| > r) \Pr(|Y-m| > r) +$$

$$\Pr(NE \mid |Y-m| \leq r) \Pr(|Y-m| \leq r)$$

$$= \Pr(NE \mid |Y-m| > r)\, 0 + \Pr(NE \mid Y=m)\, 1$$

$$= \Pr(NE \mid Y=m)$$

= Pr(no bin is empty in Exact Case when m balls are thrown)

Consequence of Two Facts

➔ Pr(some bin is still empty in Exact Case when m balls are thrown)

$= 1 - e^{-e^{-c}}$

Recall: X = # balls thrown in the exact case until every bin is non-empty

So $X > m$ occurs if and only if some bin is still empty when m balls are thrown

Thus,

$$\Pr(X > m) = 1 - e^{-e^{-c}}$$

Fact 1: Y is very close to mean

Recall:

- n = number of bins
- Y = # balls thrown in Poisson case
- $m = n \ln n + cn = E[Y]$
- $r = (2m \ln m)^{1/2}$

Fact 1: In the Poisson case, as $n \to \infty$,

$$\Pr(|Y-m| > r) = 0$$

Proof of Fact 1

First, Y is a Poisson RV with mean m

To obtain the bound for

$$\Pr(|Y-m| > r),$$

recall the Chernoff bounds: (Lecture 13, page 21)

(1) If $x > \mu$, $\Pr(Y \geq x) \leq e^{-\mu}(e\mu)^x / x^x$

(2) If $x < \mu$, $\Pr(Y \leq x) \leq e^{-\mu}(e\mu)^x / x^x$

Proof of Fact 1

So, $\Pr(|Y-m| > r) = \Pr(Y > m+r) + \Pr(Y < m-r)$

For the first term,

$$\Pr(Y > m+r) \le e^{-m} (em)^{m+r} / (m+r)^{m+r}$$

$$= e^{r} (m)^{m+r} / (m+r)^{m+r}$$

$$= \exp\{ r - (m+r) \ln ((m+r)/m) \}$$

$$= \exp\{ r - (m+r) \ln (1+ (r/m)) \}$$

Next, we use the inequality that (for $|z| < 1$)

$$\ln (1+z) \ge z - z^2/2$$

Proof of Fact 1

So, (with $r = (2m \ln m)^{1/2}$)

$\Pr(Y > m+r)$

$\leq \exp\{ r - (m+r)((r/m)-(r^2/(2m^2))) \}$

$= \exp\{ r - (m+r)((r/m)-(\ln m/m)) \}$

$= \exp\{ r - (r-\ln m) - ((r^2/m)-(r \ln m/m)) \}$

$= \exp\{ \ln m - (2 \ln m - (r \ln m/m)) \}$

$= \exp\{ -\ln m + o(\ln m) \}$

$= 0$ ... as $n \rightarrow \infty$, so that $m \rightarrow \infty$

Proof of Fact 1

On the other hand, (with $r = (2m \ln m)^{1/2}$)

$Pr(Y < m-r) \leq e^{-m} (em)^{m-r} / (m-r)^{m-r}$

$= e^{-r} (m)^{m-r} / (m-r)^{m-r}$

$= \exp\{ - r - (m-r) \ln ((m-r)/m) \}$

$\leq \exp\{ -r - (m-r)((-r/m) - (r^2/2m^2)) \}$

$= \exp\{ - r + r - r^2/(2m) - (r \ln m/m)\}$

$= \exp\{ - \ln m - o(\ln m)\}$

$= 0$ ... as $n \rightarrow \infty$, so that $m \rightarrow \infty$

Proof of Fact 1

Thus, in the Poisson case,

$0 \leq \Pr(|Y\text{-}m| > r)$

$= \Pr(Y > m+r) + \Pr(Y < m\text{-}r)$

$\leq 0 + 0$ … as $n \to \infty$, so that $m \to \infty$

$= 0$

➔ $\Pr(|Y\text{-}m| > r) = 0$ … as $n \to \infty$, so that $m \to \infty$

Fact 2

Recall:

- n = number of bins
- Y = # balls thrown in the Poisson case
- $m = n \ln n + cn = E[Y]$
- $r = (2m \ln m)^{1/2}$
- NE = the event that no bin is empty

Fact 2: In Poisson case, as $n \to \infty$,

$$\Pr(NE \mid |Y-m| \leq r) = \Pr(NE \mid Y=m)$$

Proof of Fact 2

Firstly, we observe that

$\Pr(NE \mid Y=k)$ is increasing in k … (why?)

➔ $\Pr(NE \cap Y=k) / \Pr(Y=k)$

$\leq \Pr(NE \mid Y=k+1) \leq \Pr(NE \mid Y=k+2) \leq \ldots$

In other words,

$\Pr(NE \cap Y=k) \leq \Pr(Y=k) \Pr(NE \mid Y=k+1)$

$\leq \Pr(Y=k) \Pr(NE \mid Y=k+2) \leq \ldots$

Proof of Fact 2

So, $\Pr(NE \mid |Y-m| \leq r)$

$$= \sum_{k=m-r}^{m+r} \Pr(NE \cap Y=k) \,/\, \sum_{k=m-r}^{m+r} \Pr(Y=k)$$

$$\leq \frac{\sum_{k=m-r}^{m+r} \Pr(Y=k)\,\Pr(NE \mid Y=m+r)}{\sum_{k=m-r}^{m+r} \Pr(Y=k)}$$

$$= \Pr(NE \mid Y=m+r)$$

Similarly, $\Pr(NE \mid Y=m-r) \leq \Pr(NE \mid |Y-m| \leq r)$

Proof of Fact 2

Next, we want to upper bound this term:

$$\left| \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m) \right|$$

Hopefully, we can show this to be 0

However, we don't know if $\Pr(NE \mid Y=m)$ is larger, or $\Pr(NE \mid |Y-m| \le r)$ is larger…

Let's get a bound that works for both cases

Proof of Fact 2

Case 1: Suppose Pr(NE | Y=m) is larger

Then, we know that

$$| \Pr(NE \mid |Y-m| \leq r) - \Pr(NE \mid Y=m) |$$

$$= \Pr(NE \mid Y=m) - \Pr(NE \mid |Y-m| \leq r)$$

$$\leq \Pr(NE \mid Y=m) - \Pr(NE \mid Y=m-r)$$

$$\leq \Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m-r)$$

Proof of Fact 2

Case 2: Suppose Pr(NE | Y=m) is smaller

Then, we know that

$$| \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m) |$$

$$= \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m)$$

$$\le \Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m)$$

$$\le \Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m-r)$$

Proof of Fact 2

Conclusion:

It is always true that:

$$\left| \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m) \right|$$

$$\le \Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m-r)$$

Question: What is the physical meaning of

$\Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m-r)$?

Proof of Fact 2

By Theorem on Page 7, it is the difference of the probability, in the exact case, that all bins have at least one balls when m+r balls and when m-r balls are thrown …

Also equals to Pr(success) in the following:

Step 1. Throw m-r balls

Step 2. If all bins non-empty, failure

Step 3. Else, throw 2r more balls

Step 4. If all bins non-empty, success.
Else, failure

Will Pr(success) be large? Or small?

Proof of Fact 2

Then, (with $m = n \ln n + cn$, $r = (2m \ln m)^{1/2}$)

Pr(success)

= Pr(some bins empty after m-r balls and
all bins nonempty after 2r extra balls)

≤ Pr(some bins empty after m-r balls and
a specific empty bin becomes nonempty
after 2r extra balls)

≤ Pr(a specific empty bin becomes nonempty
after 2r extra balls)

$\leq 2r/n$ [union bound] $= 0$ as $n \to \infty$

Proof of Fact 2

Thus,

$$0 \le \left| \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m) \right|$$

$$\le \Pr(NE \mid Y=m+r) - \Pr(NE \mid Y=m-r)$$

$$= \Pr(success) \le 0 \qquad \text{... as } n \to \infty$$

→ As $n \to \infty$,

$$\left| \Pr(NE \mid |Y-m| \le r) - \Pr(NE \mid Y=m) \right| = 0$$

or, $\Pr(NE \mid |Y-m| \le r) = \Pr(NE \mid Y=m)$