Probabilistic Method and Random Graphs

Lecture 5. Bins and Balls - Handling Dependency [1]

Xingwu Liu

Institute of Computing Technology
Chinese Academy of Sciences, Beijing, China

[1] The slides are mainly based on Chapter 5 of *Probability and Computing*.

Preface

Questions, comments, or suggestions?

Review: Insight into Chernoff bounds

Two questions

- Do moments uniquely determine the distribution?
- Why are Chernoff bounds so tight?

Generating functions

Invented by Abraham de Moivre to compute Fibonacci numbers.
Moment generating functions: $M_X(t) = \mathbb{E}[e^{tX}]$.
Unique when bounded or convergent around 0

Review: Large Deviation Theory

Central limit theorem: $O(\sqrt{n})$-deviation, no rate information

Chernoff bounds: large deviation, but loose

Large deviation theorem: asymptotical, tight vanishing rate

By courtesy of Cramer (1944).
Let $X_1, \cdots, X_n, \cdots \in \mathbb{R}$ be **i.i.d.** r.v. which satisfy $\mathbb{E}[e^{tX_1}] < \infty$ for $t \in \mathbb{R}$. Then for any $t > \mathbb{E}[X_1]$, we have

$$\lim_{n\to\infty} \frac{1}{n} \ln \Pr(\sum_{i=1}^{n} X_i \geq tn) = -\sup_{\lambda>0}(\lambda t - \ln \mathbb{E}[e^{\lambda X_1}]).$$

Review: Large Deviation Theory

Central limit theorem: $O(\sqrt{n})$-deviation, no rate information

Chernoff bounds: large deviation, but loose

Large deviation theorem: asymptotical, tight vanishing rate

By courtesy of Cramer (1944).
Let $X_1, \cdots, X_n, \cdots \in \mathbb{R}$ be **i.i.d.** r.v. which satisfy $\mathbb{E}[e^{tX_1}] < \infty$ for $t \in \mathbb{R}$. Then for any $t > \mathbb{E}[X_1]$, we have

$$\Pr(\sum_{i=1}^{n} X_i \geq tn) = e^{-I(t)n+o(n)}, I(t) = \sup_{\lambda>0}(\lambda t - \ln \mathbb{E}[e^{\lambda X_1}]).$$

Bins-and-Balls: Coping with Dependence

Main idea

Approximation with independence.

Focus

Approximation.

The Bins-and-Balls Model

General setting: (m, n)-model

Uniformly randomly and independently throw m balls into n bins

Extension

Multiple choice, limited capacity of bins ...

Applications

Load balancing: balls = jobs, bins = servers;
Data storage: balls = files, bins = disks;
Hashing: balls = data keys, bins = hash table slots;
Coupon Collector: balls = coupons; bins = coupon types.

Number of balls in a bin (Load of the bin) : $Bin(m, \frac{1}{n})$.

Loads of multiple bins : not independent. Why?

Application: time complexity of bucket-sort

Bucket-sort: Given $n = 2^m$ integers from $[0, 2^k)$ with $k > m$, first allocate the integers to n bins, followed by sorting each bin.
Expected time complexity: $n + \mathbb{E}[\sum_{i=1}^{n} X_i^2] = n + n\mathbb{E}[X_1^2]$.
$X_1 \sim Bin(n, \frac{1}{n})$, so $\mathbb{E}[X_1^2] = 2 - \frac{1}{n}$.

The distribution of

- Load of a certain bin
- Maximum load
- Number of bins with load r
- $\cdots$

Max. load: when does it exceed 1 w.h.p.?

The probability that max. load is 1 is

$$(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{m-1}{n}) \leq \prod_{i=1}^{m-1} e^{-\frac{i}{n}} \approx e^{-\frac{m^2}{2n}}.$$

It is less than $\frac{1}{2}$ if $m \geq \sqrt{2n\ln 2}$

Birthday paradox

$n = 365, m \geq 22.49$

Max load L: (n, n)-model

Asymptotically, $\Pr(L \geq 3\frac{\ln n}{\ln \ln n}) \leq \frac{1}{n}$

Proof

X_i: the load of bin i.
Let $k = 3\frac{\ln n}{\ln \ln n}$.
$\Pr(L \geq k) = \Pr(\bigcup_{i=1}^{n}(X_i \geq k)) \leq n \Pr(X_1 \geq k)$.
By the basic Chernoff bound on X_1 with $\mu = 1, \delta = k - 1$,

$$\begin{aligned}
\Pr(X_1 \geq k) &\leq \frac{e^{k-1}}{k^k} \leq \left(\frac{e}{k}\right)^k \\
&= \left(\frac{e \ln \ln n}{3 \ln n}\right)^{3\frac{\ln n}{\ln \ln n}} \leq \left(\frac{\ln \ln n}{\ln n}\right)^{3\frac{\ln n}{\ln \ln n}} \\
&\leq e^{(\ln \ln \ln n - \ln \ln n)\frac{3 \ln n}{\ln \ln n}} \leq \frac{1}{n^2} \text{ for big } n
\end{aligned}$$

Number of bins having load r: (m, n)-model

$r = 0$

$X_i's$ are identically distributed: $Bin(m, \frac{1}{n})$.
$\Pr(X_i = 0) = \left(1 - \frac{1}{n}\right)^m \approx e^{-\frac{m}{n}}$.
Expected number of empty bins is about $ne^{-\frac{m}{n}}$.

Load=r

$\Pr(X_i = r) = \binom{m}{r}\frac{1}{n^r}\left(1 - \frac{1}{n}\right)^{m-r}$.
When $r \ll \min\{m, n\}$, $\Pr(X_i = r) \approx e^{-\frac{m}{n}}\frac{\left(\frac{m}{n}\right)^r}{r!}$.
Expected number of load-r bins is about $ne^{-\frac{m}{n}}\frac{\left(\frac{m}{n}\right)^r}{r!}$.

Poisson distribution

$\sum_r e^{-\mu}\frac{\mu^r}{r!} = 1$ due to $e^x = \sum_r \frac{x^r}{r!}$.
Nonnegative-integer-valued r.v. X_μ: $\Pr(X_\mu = r) = e^{-\mu}\frac{\mu^r}{r!}$.

Basic Properties of Poisson distribution

Low-order moments

$\mathbb{E}[X_\mu] = Var[X_\mu] = \mu$.

Moment generation function

$$M_{X_\mu}(t) = \mathbb{E}[e^{tX_\mu}] = \sum_r \frac{e^{-\mu}\mu^r}{r!}e^{tr} = e^{-\mu}\sum_r \frac{\left(\mu e^t\right)^r}{r!} = e^{\mu(e^t-1)}.$$

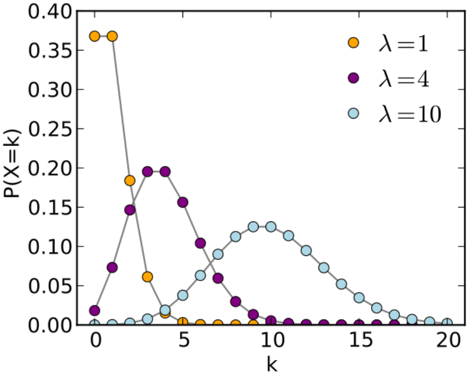
Additive

By uniqueness of moment generation functions,
$X_{\mu_1} + X_{\mu_2} = X_{\mu_1+\mu_2}$ if independent.

Chernoff-like bounds

1. If $x > \mu$, then $\Pr(X_\mu \geq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}$.
2. If $x < \mu$, then $\Pr(X_\mu \leq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}$.

Graphs of Poisson Density Functions



Applications and Story

Occurrences of rare events during a fixed interval

- Typos per page in printed books.
- Number of bomb hits per $0.25km^2$ in South London during World War II.
- The number of goals in sports involving two competing teams.
- *The number of soldiers killed by horse-kicks each year in Prussian cavalry corps in the (late) 19th century.*

Story of Poisson distribution

1837, Poisson, *Research on the Probability of Judgments in Criminal and Civil Matters*.
Appeared in 1711, de Moivre. (Stigler's law of eponymy, 1980)
First practical application (next page)

First practical application of Poisson distribution

Reliability engineering: Ladislaus Bortkiewicz (1868-1931)

- Russian economist and statistician of Polish ancestry, mostly lived in Germany
- Known for Poisson Dis. and Marxian econ.
- The book *The Law of Small Numbers*, 1898

- Annual Horse-kick data of 14 cavalry corps over 20 years
- Events with low probability in a large population follow a Poisson distribution

No. deaths k	Freq.	Poisson approx. $200 \times \mathbb{P}(\text{Poi}(0.61) = k)$
0	109	108.67
1	65	66.29
2	22	20.22
3	3	4.11
4	1	0.63
5	0	0.08
6	0	0.01

Law of Small Numbers (Poisson Convergence)

Poisson convergence of binomial distribution

Assume that $X_n \sim Bin(n, p_n)$ with $\lim_{n\to\infty} np_n = \lambda$. For any fixed k, $\lim_{n\to\infty} \Pr(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

It is intuitively acceptable (by their figures)

It can be used to approximately calculate Binomial distribution $Bin(n, p)$, but take care.
$n > 100, p < 0.01, np < 20$.

Error bounds implies the convergence

$$e^{\frac{p(k-np)}{1-p} - \frac{k(k-1)}{2(n-k+1)}} \leq \frac{\Pr(Bin(n,p)=k)}{\Pr(Poi(np)=k)} \leq e^{kp - \frac{k(k-1)}{2n}}.$$

Proof of the error bounds

Error bounds

$$e^{\frac{p(k-np)}{1-p}-\frac{k(k-1)}{2(n-k+1)}} \le \frac{\Pr(Bin(n,p)=k)}{\Pr(Poi(np)=k)} \le e^{kp-\frac{k(k-1)}{2n}}.$$

Proof

$A_{n,p,k} \triangleq \frac{\Pr(Bin(n,p)=k)}{\Pr(Poi(np)=k)} = \prod_{j=1}^{k-1}\left(1-\frac{j}{n}\right)e^{np}(1-p)^{n-k}$ for $0 \le k \le n$ and it's 0 otherwise.

Upper bound

$$A_{n,p,k} \le e^{-\sum_{j=1}^{k-1}\frac{j}{n}+np-(n-k)p} = e^{kp-\frac{k(k-1)}{2n}}.$$

Lower bound

$$\begin{aligned} A_{n,p,k} &\ge e^{-\sum_{j=1}^{k-1}\frac{j/n}{1-j/n}+np-(n-k)\frac{p}{1-p}} \\ &= e^{-\sum_{j=1}^{k-1}\frac{j}{n-j}-\frac{p(np-k)}{1-p}} \ge e^{\frac{p(k-np)}{1-p}-\frac{k(k-1)}{2(n-k+1)}}. \end{aligned}$$

Generalize LSN to weak dependence

Poisson convergence with weak dependence

For each n, let Y_n be # occurrences of events $B_1^n, \dots B_n^n$. If

- $\lim_{n\to\infty} \mathbb{E}[Y_n] = \lambda$
- For any k, $\lim_{n\to\infty} \sum_{1\le i_1<\dots<i_k\le n} \Pr(\bigcap_{r=1}^k B_{i_r}^n) = \frac{\lambda^k}{k!}$,

then $Y_n \to Poi(\lambda)$, i.e. $\Pr(Y_n = j) \to \frac{e^{-\lambda}\lambda^j}{j!}$ for any $j \ge 0$

Basic idea of the proof for $j = 0$:
Use Taylor series of $e^{-\lambda}$ and Bonferroni inequalities

- $\Pr(\bigcup_{i\ge 1}^{n} B_i^n) \le \sum_{l=1}^{r} (-1)^{l-1} \sum_{i_1<i_2<\dots<i_l} \Pr(\bigcap_{r=1}^{l} B_{i_r}^n)$ for odd r
- $\Pr(\bigcup_{i\ge 1}^{n} B_i^n) \ge \sum_{l=1}^{r} (-1)^{l-1} \sum_{i_1<i_2<\dots<i_l} \Pr(\bigcap_{r=1}^{l} B_{i_r}^n)$ for even r

Remarks on the case of weak dependence

Intuitive explanation

If X is $\#$ occurrences of a large collection of nearly independent rare events, the $X \sim Poi(\mathbb{E}[X])$

Application

- The number of people who get their own hats back after a random permutation of the hats
- The number of pairs having the same birthday
- The number of isolated vertices in random graph $G(n, \frac{\ln n + c}{n})$

It can be further generalized

Generalize LSN to strong dependence

Poisson convergence with strong dependence, 1975

Stein-Chen Theorem: If Y_n is the sum of Poisson trials $X_1, \ldots X_n$ and $\lambda = \mathbb{E}[Y_n]$, then for any $A \subseteq \mathbb{Z}_+$,

$$|\Pr(Y_n \in A) - \Pr(Poi(\lambda) \in A)| \leq \min\left\{1, \frac{1}{\lambda}\right\} \sum_{i=1}^{n} p_i \mathbb{E}[|U_i - V_i|].$$

where $U_i \sim Y_n, 1 + V_i \sim Y_n | X_i = 1$, $p_i = \Pr(X_i = 1)$.

Intuitive explanation

Poisson approximation remains valid even if the Bernoulli r.v.s are strongly dependent and have different expectations.

Remarks on the law of small numbers

Law of small numbers vs Law of large numbers (CLT)

- Poisson approximation vs Normal approximation
- Small number vs arbitrary number
- Sums of different sets vs partial sums of one sequence

Relation between Poisson and Normal distribution

Should be related since both approximate binomial distribution.
When $\lambda \to \infty$, Poisson converges to Normal.
Specifically, $\lim_{\lambda\to\infty} \sum_{\alpha<k<\beta} \frac{\lambda^k e^{-\lambda}}{k!} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx$.
Where $a = (\alpha-\lambda)/\sqrt{\lambda}, b = (\beta-\lambda)/\sqrt{\lambda}$ are fixed.

Intuitive argument

Uniqueness+continuity of moment generating functions.

References

1. https://www.math.illinois.edu/~psdey/414CourseNotes.pdf
2. http://willperkins.org/6221/slides/poisson.pdf