Probabilistic Method and Random Graphs

Lecture 4. Large Deviation Theory & Bins and Balls[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?

A brief review of Lecture 3

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: why?

Chernoff bound in a big picture

Fundamental laws of probability theory

Law of large numbers (Cardano, Jacob Bernoulli 1713, Poisson 1837): The sample average converges to the expected value.

Central limit theorem (Abraham de Moivre 1733, Laplace 1812, Lyapunov 1901, Pólya 1920): The arithmetic mean of independent random variables is approximately normally distributed.

$$\lim_{n\to\infty} \Pr\left(\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \leq x\right) = \Phi\left(\frac{x}{\sigma}\right)$$

Marvelous but ...

Say nothing about the rate of convergence

Large deviation theory

How fast does it converge? Beyond central limit theorem

Motivation

X_n: the number of heads in n flips of a fair coin.
By the central limit theorem, $\Pr(X_n \geq \frac{n}{2} + \sqrt{n}) \to 1 - \Phi(1)$.
What about $\Pr(X_n \geq \frac{n}{2} + \frac{n}{3})$? Nothing but converging to 0.

Chernoff bounds say...

$$\Pr(X_n \geq \tfrac{n}{2} + \tfrac{n}{3}) \leq \left(\frac{e^{\frac{2}{3}}}{\left(\frac{5}{3}\right)^{\frac{5}{3}}} \right)^{\frac{n}{2}} \approx e^{-0.092n}.$$

Actually

Direct calculation shows that
$\Pr(X_n \geq \frac{n}{2} + \frac{n}{3}) \approx e^{-0.2426n + o(n)} \ll \text{Chernoff bound.}$

Oh, no!

Mission of Large Deviation Theory

Find the asymptotic probabilities of *rare* events - how do they decay to 0 as $n \to \infty$?

Rare events mean large deviation.
So large that CLT is almost useless (deviation up to $\sqrt{n}$).

Intuition

Inspired by Chernoff bounds, conjecture that probabilities of rare events will be exponentially small in n : e^{-cn} for some c.
Q: Does $\lim_{n\to\infty} \frac{1}{n} \ln \Pr(\mathcal{E}_n^{\text{rare}})$ exist? If so, what's it?

Large Deviation Principle

Simple form (By courtesy of Cramer, 1938)

Let $X_1, \ldots X_n, \ldots \in \mathbb{R}$ 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

$$\lim_{n\to\infty} \frac{1}{n} \ln \Pr(\sum_{i=1}^{n} X_i \geq tn) = -I(t),$$

where

$$I(t) \triangleq \sup_{\lambda>0} \lambda t - \ln \mathbb{E}[e^{\lambda X_1}].$$

Remark

$I(\cdot)$: rate function.
Many variants: the factor $\frac{1}{n}$, tn in the events, random variables

Large Deviation Principle: Proof

Large Deviation Principle

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

Proof: Upper bound

Let $Y_n = \frac{\sum_{i=1}^{n} X_i}{n}$, $M(\lambda) = \mathbb{E}[e^{\lambda X_1}]$, and $\psi(\lambda) = \ln M(\lambda)$.

$\Pr(Y_n \geq t) \leq e^{-\lambda n t}(M(\lambda))^n$ for any $\lambda \geq 0$.

$\frac{1}{n} \ln \Pr(Y_n \geq t) \leq -\lambda t + \psi(\lambda)$.

$\frac{1}{n} \ln \Pr(Y_n \geq t) \leq -\sup_{\lambda \geq 0}(\lambda t - \psi(\lambda))$.

Large Deviation Principle: Proof

Lower bound

The maximizer λ_0 of $\lambda t - \psi(\lambda)$ satisfies $t = \int \frac{x e^{\lambda_0 x}}{M(\lambda_0)} d\mu(x)$.

Let $d\mu_0(x) = \frac{e^{\lambda_0 x}}{M(\lambda_0)} d\mu(x)$. Its expectation $\int x d\mu_0(x) = t$.

Let $A = \{Y_n \geq t\} \subseteq \mathbb{R}^n$, $A_\delta = \{Y_n \in [t, t+\delta]\} \subseteq \mathbb{R}^n$.

$$
\begin{aligned}
\Pr_\mu(A) \geq \Pr_\mu(A_\delta) &= \int_{A_\delta} \Pi_{i=1}^n d\mu(x_i) \\
&= \int_{A_\delta} (M(\lambda_0))^n e^{-\lambda_0 \sum_{i=1}^n x_i} \Pi_{i=1}^n d\mu_0(x_i) \\
&\geq (M(\lambda_0) e^{-\lambda_0 (t+\delta)})^n \Pr_{\mu_0}(A_\delta).
\end{aligned}
$$

Applying CLT to μ_0, we have $\lim_{n\to\infty} \Pr_{\mu_0}(A_\delta) = \frac{1}{2}$.

$\lim_{n\to\infty} \frac{1}{n} \ln \Pr(Y_n \geq t) \geq \psi(\lambda_0) - (t+\delta)\lambda_0$, and let $\delta \to 0$.

Remarks

Large deviation theory vs CLT

Seemingly easy to get exponential decay in many cases, but hard to calculate.

Chernoff bounds concern large deviation

- Con: Generally weaker
- Pro: Always holds, not just asymptotically

Key assumption

Independence!

Bins-and-Balls: Coping with Dependence

Main idea

Approximation with independence.

Focus

Approximation.

The Bins-and-Balls Model

General setting: (m, n)-model

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.

Basic Properties

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

Numbers of balls in 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 $\mathbb{E}[X_1^2] = 2 - \frac{1}{n}$.

Topics of Bins-and-Balls Model

The distribution of

Number of balls in a certain bin

Maximum load

Number of bins containing r balls

...

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

$n = 365, m \geq 22.49$

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

Proof

X_i: the number of balls in bin i.
$\Pr(X_1 \geq k) \leq \binom{n}{k}\frac{1}{n^k} \leq \frac{1}{k!}$.
$\frac{k^k}{k!} < \sum_i \frac{k^i}{i!} = e^k \Rightarrow \frac{1}{k!} \leq \left(\frac{e}{k}\right)^k$.

$$\begin{aligned}\Pr\left(L \geq 3\frac{\ln n}{\ln \ln n}\right) &\leq n\left(\frac{e \ln \ln n}{3 \ln n}\right)^{3\frac{\ln n}{\ln \ln n}} \\ &\leq n\left(\frac{\ln \ln n}{\ln n}\right)^{3\frac{\ln n}{\ln \ln n}} \\ &\leq e^{\ln n + (\ln \ln \ln n - \ln \ln n)\frac{3 \ln n}{\ln \ln n}} \leq \frac{1}{n}.\end{aligned}$$

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

$r = 0$

The distribution of $X_i's$ are identical: $Bin(m, \frac{1}{n})$.
$\Pr(X_i = 0) = \left(1 - \frac{1}{n}\right)^m \approx e^{-\frac{m}{n}}$.
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

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

Basic Properties of Poisson distribution

Low-order moments

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

Moment generation function

$M_{X_\mu}(t) = \mathbb{E}[e^{tX_\mu}] = \sum_k \frac{e^{-\mu}\mu^k}{k!}e^{tk} = e^{\mu(e^t-1)}.$

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 $\Pr(X_\mu \leq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}$.

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
  - Famous for Poisson distribution and Marxian economics
- 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

References

1. http://willperkins.org/6221/slides/largedeviations.pdf