Lecture 2: Concentration Inequalities

Nicolás Rivera
John Sylvester   Luca Zanetti   Thomas Sauerwald

Lent 1920

UNIVERSITY OF CAMBRIDGE

Concentration Inequalities

- Concentration refers to the phenomena where random variables are very close to their mean.
- This is very useful in randomised algorithms as it ensures an **almost** deterministic behaviour
- It gives us the best of two worlds:
  1. **Randomised Algorithms:** Easy to Design and Implement
  2. **Deterministic Algorithms:** They do what they claim

Outline

Basic Concentration inequalities

Boole Inequality - Union Bound

Boole Inequality is the simplest of all probabilistic inequalities but it is quite handy.

Union-Bound

Let $A_1, \ldots, A_n$ be a collection of events in Σ. Then

$$\mathbf{P}\left[\bigcup_{i=1}^{n} A_i\right] \le \sum_{i=1}^{n} \mathbf{P}[\,A_i\,]$$

For a short proof:

1. Denote by $\mathbf{1}_{A_i}$ the random variable that takes value 1 if A_i holds, 0 otherwise[1]
2. $\mathbf{E}[\,\mathbf{1}_{A_i}\,] = \mathbf{P}[\,A_i\,]$ (**Check this**)
3. It is clear that $\mathbf{1}_{\bigcup_{i=1}^{n} A_i} \le \sum_{i=1}^{n} \mathbf{1}_{A_i}$ (**Check this**)
4. Taking expectation we conclude the result.

[1] Formally, this random variable is $\mathbf{1}_{A_i}(\omega)$ which takes value 1 if $\omega \in A_i$, 0 otherwise

Example: Coupon Collector

This is a very important example in the design and analysis of randomised algorithms.

Coupon Collector Problem

> Suppose that there are N coupons to be collected from the cereal box. Every morning you open your cereal box and get one coupon. Each coupon appears with the same probability in the box, i.e. $1/N$.

Exercise

1. Prove it takes $N\sum_{k=1}^{N}\frac{1}{k} \approx N\log N$ expected boxes to collect all coupons
2. Use the Union Bound to prove that the probability it takes us more than $N\log(N) + cN$ boxes to collect all the coupons is smaller than or equal to e^{-c}.

- **It is useful to remember that** $1 - x \leq e^{-x}$ **for all** $x < 1$

Markov Inequality

Markov Inequality is the second most basic, and probably the most important inequality in probability.

Markov Inequality

If X is a non-negative random variable and $a > 0$, then

$$\mathbf{P}[X \geq a] \leq \mathbf{E}[X]/a.$$

Again, we can write a short proof:

- $\mathbf{P}[X \geq a] = \mathbf{E}\left[\mathbf{1}_{\{X \geq a\}}\right]$
- $\mathbf{1}_{\{X \geq a\}} \leq \frac{X}{a}\mathbf{1}_{\{X \geq a\}} \leq \frac{X}{a}$
- Then $\mathbf{P}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$

Markov Inequality

If X is a non-negative random variable and $a > 0$, then

$$\mathbf{P}[\, X \geq a \,] \leq \mathbf{E}[\, X \,] \,/ a.$$

From the Markov Inequality, we get the following

Chevyshev Inequality

If X is a random variable and $a > 0$, then

$$\mathbf{P}[\, |X - \mathbf{E}[\, X \,]| \geq a \,] \leq \mathbf{Var}\,[\, X \,] \,/ a^2.$$

Exercise: Prove the Chevyshev Inequality

Why are these inequalities important?

Example: Coin flip

Consider n fair coins and let X be the total number of head. In an experiment we expect to see around $n/2$ heads. Can we justify that? Let $\delta > 0$

- Markov inequality :

$$\mathbf{P}[\, X \geq (1+\delta)(n/2)\,] \leq \frac{\mathbf{E}[\, X\,]}{(1+\delta)(n/2)} = \frac{n/2}{(1+\delta)(n/2)} = \frac{1}{1+\delta}$$

Not good! Independent of n

- Chebychev inequality :

$$\begin{aligned} \mathbf{P}[\,(X - n/2) \geq \delta(n/2)\,] &\leq \mathbf{P}[\,|X - n/2| \geq \delta(n/2)\,] \\ &\leq \frac{4\mathbf{Var}[\,X\,]}{\delta^2 n^2} = \frac{1}{\delta^2 n} \end{aligned}$$

Better! Linear in n

Example: RandMaxCut

Recall the Max-Cut problem from last class.

- Given a undirected graph $G = (V, E)$.
- Denote by $e(A, A^c)$ the number of edges with one end-point in A and the other in A^c.
- We want to find $\max_{A \subseteq V} e(A, A^c)$

Algorithm: RandMaxCut

Input $G = (V, E)$.
-Start with $S = \emptyset$.
-For each $v \in V$ add v to S independently with probability $1/2$.
Return S.

Example: RandMaxCut

For the analysis it is useful to define a few random variables

- Let A_i the event that i belongs to the random set S
- $X_i = \mathbf{1}_{A_i}$, is an **indicator random variable**, that takes value 1 if A_i holds, 0 otherwise
- Define $Y_{ij} = X_i(1 - X_j) + (1 - X_i)X_j$
- We are interested in the random variable $Z = e(S, S^c) = \sum_{\{i,j\} \in E(G)} Y_{ij}$
- $\mathbf{E}[\, Z \,] = \sum_{\{i,j\} \in E(G)} \mathbf{E}[\, Y_{ij} \,] = |E|/2$
- Observe that

$$\frac{\mathbf{E}[\, Z \,]}{\max_{U \subseteq V} e(U, U^c)} \geq \frac{\mathbf{E}[\, Z \,]}{|E|} \geq 1/2$$

- Hence, in expectation, our algorithm is a $(1/2)$-approximation. I.e. the expected value of our solution is at least half of the optimal solution.

Example: RandMaxCut

Lets compute the variance of Z.

$$\begin{aligned}
\mathbf{E}\left[Z^2\right] &= \sum_{ij\in E}\sum_{k\ell\in E}\mathbf{E}[\,Y_{ij}Y_{k\ell}\,] && (1)\\
&= \sum_{ij\in E}\mathbf{E}[\,Y_{ij}\,] + \sum_{ij\in E}\left(\sum_{\substack{k\ell\in E\\ k\ell\neq ij}}\mathbf{E}[\,Y_{ij}Y_{k\ell}\,]\right) && (2)\\
&\leq \sum_{ij\in E}\mathbf{E}[\,Y_{ij}\,] + \sum_{ij\in E}\left(\sum_{\substack{k\ell\in E\\ k\ell\neq ij}}\mathbf{E}[\,Y_{ij}\,]\,\mathbf{E}[\,Y_{k\ell}\,]\right) && (3)\\
&\leq \sum_{ij\in E}\mathbf{E}[\,Y_{ij}\,] + \sum_{ij\in E}\left(\sum_{k\ell\in E}\mathbf{E}[\,Y_{ij}\,]\,\mathbf{E}[\,Y_{k\ell}\,]\right) && (4)\\
&= \mathbf{E}[\,Z\,] + \mathbf{E}[\,Z\,]^2 && (5)
\end{aligned}$$

Using that $\mathbf{Var}\,[\,Z\,] = \mathbf{E}\left[Z^2\right] - \mathbf{E}[\,Z\,]^2$, we get

$$\mathbf{Var}\,[\,Z\,] \leq \mathbf{E}[\,Z\,].$$

Example: RandMaxCut

- We already know that $\mathbf{E}[\, e(S, S^c)\,] = |E|/2$
- Also $\mathbf{Var}\,[\, e(S, S^c)\,] \leq \mathbf{E}[\, e(S, S^c)\,]$.
- Chevyshev's inequality says

$$\mathbf{P}\left[\, |e(S, S^c) - \mathbf{E}\left[\, e(S, S^c)\,\right]| \geq a\,\right] \leq \frac{\mathbf{Var}\,[\, e(S, S^c)\,]}{a^2} \leq \frac{\mathbf{E}[\, e(S, S^c)\,]}{a^2}$$

- Choose $a = C\sqrt{|E|}$, where C can be a large constant. Then

$$\mathbf{P}\left[\, |e(S, S^c) - \mathbf{E}\left[\, e(S, S^c)\,\right]| \geq C\sqrt{|E|}\,\right] \leq \frac{1}{2C^2}$$

- This means that the random solution does not moves much from its mean.

Markov and Chebychev use the first and second moment of the random variable. Can we keep going?

- Yes.

We can consider first, second, third and more moments! that is the basic idea behind the Chernoff Bounds

Chernoff Bounds

Chernoff Bounds

Suppose $X_1, \dots, X_n$ are independent Bernoulli random variables with parameter p_i. Let $X = X_1 + \ldots + X_n$ and $\mu = \mathbf{E}[\, X\,] = \sum p_i$. Then, for any $\delta > 0$ it holds that

$$\mathbf{P}[\, X \geq (1+\delta)\mu\,] \leq \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}.$$

and for $t > \mu$ it holds that

$$\mathbf{P}[\, X \geq t\,] \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^{t},$$

Example: Coin Flip

Consider n fair coins and let X be the total number of head. Then

$$\mathbf{P}[\, X \geq (1+\delta)(n/2) \,] \leq \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right]^{n/2}$$

- Note that the above expression equals 1 only for $\delta = 0$, and then it gives a value strictly less than 1 (**Ex: check this!**)
- Note as well the inequality is exponential in n, (for fixed δ) i.e. much better than Chebychev inequality.

Example: Coin Flip

Consider 100 independent coin flips . We wish to find an upper bound on the probability that the number of heads is greater or equal than 75.

- Markov's inequality : $X = \sum_{i=1}^{100} X_i$, $X_i \in \{0,1\}$ and $\mathbf{E}[X] = 100 \cdot \frac{1}{2} = 50$.

$$\mathbf{P}[X \geq 3/2 \cdot \mathbf{E}[X]] \leq 2/3 = 0.666.$$

- Chebyshev's inequality : $\mathbf{Var}[X] = \sum_{i=1}^{100} \mathbf{Var}[X_i] = 100 \cdot (1/2)^2 = 25$.

$$\mathbf{P}[|X - \mu| \geq t] \leq \frac{\mathbf{Var}[X]}{t^2},$$

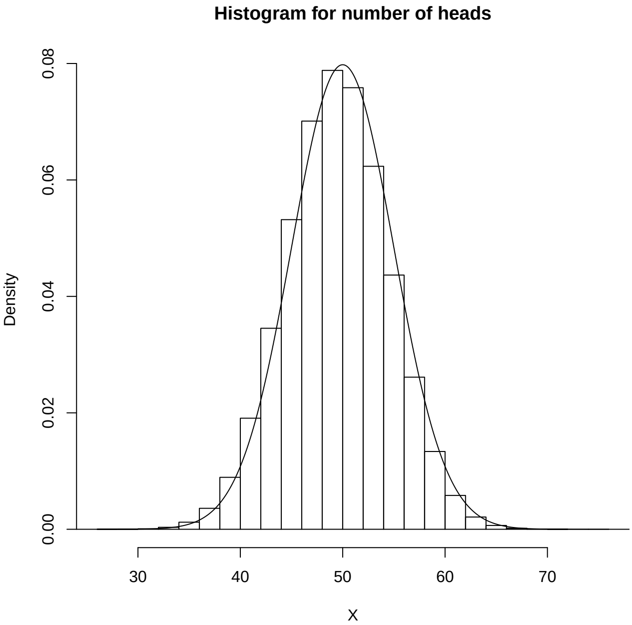
  and plugging in $t = 25$ gives an upper bound of $25/25^2 = 1/25 = 0.04$, much better than what we obtained by Markov's inequality.

- The Chernoff bound (first) with $\delta = 1/2$ gives:

$$\mathbf{P}[X \geq 3/2 \cdot \mathbf{E}[X]] \leq \left(\frac{e^{1/2}}{(3/2)^{3/2}} \right)^{50} = 0.004472$$

- the exact probability is $0.00000028\ldots$

Chernoff bound yields a more accurate result but needs independence!


Histogram for number of heads
Density
0.00
0.02
0.04
0.06
0.08
30
40
50
60
70
X