COMPSCI 630: Randomized Algorithms February 5, 2018

Lecture 7

Lecturer: Rong Ge *Scribe: Shweta Patwa*

1 Overview

In this lecture, we introduce concentration bounds. We look at the (multiplicative) Chernoff bound, and give a proof for the bound. We also discuss the additive form of Chernoff bound, i.e., the Chernoff-Hoeffding theorem, and the Bernstein's inequality.

2 Introduction

We investigate the phenomenon of concentration of the probability of a random variable X around its expectation $E[X]$. More precisely, we want $Pr[|X - E[X]| > t]$ to be upper bounded by a small quantity. In some cases, we may want to use the median instead of the expecation, which happen to be close for most of the cases we will encounter.

Theorem 1. *(Central Limit Theorem) Given $X_1, X_2, \ldots, X_n$ i.i.d random variables with $E[X_i] = 0$ and $E[X_i^2] = 1$, for all $i = 1, 2, \ldots, n$. Define $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Then,*

$$\lim_{n \to \infty} S_n \xrightarrow{d} N(0,1)$$

, where d is the following distance measure: given X, Y distributions over $\mathbb{R}$, $d(X,Y) = \sup_{\tau \in \mathbb{R}} |Pr[X \leq \tau] - Pr[Y \leq \tau]|$, i.e., the maximum difference between the two cdfs.

Thus, $\lim_{n \to \infty} d(S_n, N(0,1)) = 0$. For variational distance, δ, $\lim_{n \to \infty} \delta(S_n, N(0,1))$ can be as large as 1 and does not converge because S_n is discrete for $X_i = \pm 1$, whereas the Gaussian is continuous; follows from the definition of variational distance.

Here are some issues we run into:

1. Asymptotic result ($n \to \infty$).
   We want a rate of convergence.
   "Bery-Essen" theorem - more quantitative version of the Central Limit Theorem.

2. It does not capture the tail behavior very well. Tail is defined to be far away from the expectation, say, $\sqrt{10}\sigma$ from the expectation.

3 Chernoff bound (multiplicative version)

3.1 Motivation

Recall from last time that Chebyshev's inequality on a r.v. X gives us $Pr[|X-E[X]| \geq t] \leq \frac{Var[X]}{t^2}$. Suppose we toss a fair coin n times, and X is the r.v. representing the number of times heads shows up. Then, $E[X] = n/2$ and $Var[X] = n/4$. We can apply Chebyshev's for the following values of t:

- $t = \sqrt{n}$: $Pr[|X-E[X]| \geq t] \leq \frac{n/4}{n} = \frac{1}{4}$.
- $t = n/2$: $Pr[|X-E[X]| \geq t] \leq \frac{1}{n}$. But this is a lose upper bound.

3.2 Chernoff bound and proof

Theorem 2. *Let $X_1, X_2, \ldots, X_n$ be independent Bernouilli $(0/1)$ r.v.s with $Pr[X_i = 1] = p_i$. Define $X := \sum_i X_i$, so $\mu = E[X] = \sum_i p_i$. Then,*

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

Examples:

1. $p_i = 1/2$ and
   $\delta = 1$ or any other constant: the RHS of the inequality equals $(\frac{e}{4})^{n/2} = e^{-\Omega(n)}$ (exponential tail).
   $\delta = \varepsilon$: RHS $\approx e^{-\Omega(\varepsilon^2 n)}$. Note that for $\varepsilon = 1/\sqrt{n}$ we still get an exponential tail.
   Therefore, it is tighter than Chebyshev when far from expectation.

2. Consider the following r.v.

$$Y = \begin{cases} X & \text{,w.p. } 1-1/n \\ \frac{n}{2}+\frac{n}{2} & \text{,w.p. } 1/n \end{cases}$$

   with $Var[Y] = O(n)$. Then, Chebyshev cannot distinguish between X and Y. Concentration is very different. Y does not have "concentration", and could be far away from its expectation.

Definition 1. *(Moment generating function, also referred to as Laplace transform) Consider a r.v. X. $E[X], E[X^2]$ can be computed. Even $E[X^k]$, i.e., the k^{th} "moment" of the distribution can be computed using the following $(t > 0)$:*

$$E[e^{tX}] = \sum_{k=0}^{\infty} \frac{t^k E[X^k]}{k!} \qquad \textit{(from Taylor's expansion and linearity of expectation)}$$

, where $E[X^k]$ is the k^{th} moment and is the coefficient of $\frac{t^k}{k!}$ in the expansion.

Intuition: X larger than $E[X]$, then e^{tX} is much larger than $e^{tE[X]}$.

Proof. (Theorem 2)

$$\begin{aligned} Pr[X > (1+\delta)\mu] &= Pr[e^{tX} > e^{t(1+\delta)\mu}] \\ &< \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} \qquad \text{(by Markov's inequality)} \end{aligned}$$

The bound holds $\forall t > 0$.

$$
\begin{aligned}
E[e^{tX}] &= E[e^{t(X_1+X_2+\ldots+X_n)}] \\
&= \prod_i E[e^{tX_i}] && (X_i's \text{ are independent}) \\
&= \prod_i (p_i(e^t-1)+1) \\
&\le \prod_i e^{p_i(e^t-1)} \\
&= e^{(e^t-1)\mu}
\end{aligned}
$$

Thus,

$$Pr[X > (1+\delta)\mu] < \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} \le \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}$$

Set $t = \log(1+\delta)$, $\frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}$ simplifies to $\left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$. □

4 Chernoff-Hoeffding theorem

Here, we look at the additive form of the Chernoff bound.
Let $p = \frac{\mu}{n}$, i.e., the average probability in some sense.

$$
\begin{aligned}
&Pr[\frac{1}{n}X \ge p+\varepsilon] \le e^{-KL(p+\varepsilon||p)n} \le e^{-2e^2 n} && (KL(p+\varepsilon||p) \ge 2\varepsilon^2 \quad ,\forall p) \\
&Pr[\frac{1}{n}X \ge p-\varepsilon] \le e^{-KL(p-\varepsilon||p)n}
\end{aligned}
$$

5 Bernstein's inequality

Let $X_1, \ldots, X_n$ be independent r.v.s with $|X_i| \le R$, $E[X_i] = 0$ and variance $\sigma^2 = \sum_{i=1}^{n} E[X_i^2]$. Then,

$$Pr[\sum_{i=1}^{n} X_i \ge t] \le exp(\frac{-t^2/2}{\sigma^2 + Rt/3}).$$

We often want the RHS to be at most $1/poly(n)$ say $1/n^{10}$. It suffices to choose $t = \Theta(\sigma\sqrt{\log n} + R\log n)$.
Random variables can be continuous, unlike discrete for Chernoff bounds.

Reasons why concentration could fail:

1. Variance is large, so r.v.s would not concentrate.

2. $Pr[|X_i| \text{ is large}]$ is not small enough.

6 Summary

We motivate the idea of concentration of the probability of a random variable around its expectation, and study the following bounds: Chernoff bound (additive and multiplicative) and Bernstein's inequality.