IE613: Online Machine Learning **Jan-Apr 2018**

Lecture 18: Concentration Inequalities

Lecturer: M. K. Hanawal *Scribes: Piyush Mittal*

18.1 Introduction

The laws of large numbers of classical probability theory state that sums of independent random variables are, under very mild conditions, close to their expectation with a large probability. Such sums are the most basic examples of random variables concentrated around their mean. The purpose of these notes is to give an introduction to some of these general concentration inequalities.
The inequalities discussed in these notes bound tail probabilities of general functions of independent random variables.

18.2 Basics

To make these notes self-contained, we first briefly introduce some of the basic inequalities of probability theory.

18.2.1 Markov Inequality

First of all, let X be a non-negative random variable. Then we know that,

$$\begin{aligned} \mathbb{E}\left[X\right] &= \int_0^\infty P\left(X \geq t\right) dt \\ \Rightarrow \mathbb{E}\left[X\right] &\geq \int_0^\tau P\left(X \geq t\right) dt \\ \Rightarrow \mathbb{E}\left[X\right] &\geq P\left(X \geq \tau\right) \tau \\ \Rightarrow P\left(X \geq \tau\right) &\leq \frac{\mathbb{E}\left[X\right]}{\tau} \end{aligned} \tag{Markov-Inequality}$$

18.2.2 Chebyshev Inequality

If ϕ is a non-decreasing non-negative function, then applying Markov-Inequality,

$$P\left(X \geq \tau\right) = P\left(\phi\left(X\right) \geq \phi\left(\tau\right)\right) \leq \frac{\mathbb{E}\left[\phi\left(X\right)\right]}{\phi\left(\tau\right)}$$

Let $\phi(X) = |X - \mathbb{E}[X]|^2$ and the constant $= \tau^2$,

$$\Rightarrow \; P\left(|X - \mathbb{E}[X]|^2 \geq \tau^2\right) \; \leq \; \frac{\mathbb{E}\left[|X - \mathbb{E}[X]|^2\right]}{\tau^2} \; = \; \frac{Var(X)}{\tau^2}$$

$$\Rightarrow \; P\left(|X - \mathbb{E}[X]| \geq \tau\right) \; = \; P\left(|X - \mathbb{E}[X]|^2 \geq \tau^2\right) \; \leq \; \frac{Var(X)}{\tau^2}$$

(Chebyshev's Inequality)

If $\phi(X) = |X - \mathbb{E}[X]|^q$ and the constant $= \tau^q$

$$\Rightarrow \; P\left(|X - \mathbb{E}[X]|^q \geq \tau^q\right) \; \leq \; \frac{\mathbb{E}\left[|X - \mathbb{E}[X]|^q\right]}{\tau^q}$$

For any non-decreasing fuction ϕ,

$$\Rightarrow \; P\left(|X - \mathbb{E}[X]| \geq \tau\right) \; = \; P\left(|X - \mathbb{E}[X]|^q \geq \tau^q\right) \; \leq \; \frac{\mathbb{E}\left[|X - \mathbb{E}[X]|^q\right]}{\tau^q} \qquad \forall \; q$$

In specific examples one may choose the value of q to optimize the obtained upper bound. Such moment bounds often provide with very sharp estimates of the tail probabilities.

$$\Rightarrow \; P\left(|X - \mathbb{E}[X]| \geq \tau\right) \; \leq \; \min_q \left[\frac{\mathbb{E}\left[|X - \mathbb{E}[X]|^q\right]}{\tau^q}\right]$$

18.2.3 Chernoff's bound

Let $\tau \in \mathbb{R}$, X is bounded and $S \in \mathbb{R}^+$. Then,

$$P(X \geq \tau) \; = \; P(SX \geq S\tau) \; = \; P\left(e^{SX} \geq e^{S\tau}\right)$$

Applying Markov Inequality:

$$\Rightarrow \; P\left(e^{SX} \geq e^{S\tau}\right) \; \leq \; \frac{\mathbb{E}\left[e^{SX}\right]}{e^{S\tau}}$$

$$\Rightarrow \; P(X \geq \tau) \; \leq \; \frac{\mathbb{E}\left[e^{SX}\right]}{e^{S\tau}} \qquad \forall \; S$$

In Chernoff's method, we find an $S > 0$ that minimizes the upper bound or makes the upper bound small.

$$\Rightarrow \; P(X \geq \tau) \; \leq \; \min_S \left[\frac{\mathbb{E}\left[e^{SX}\right]}{e^{S\tau}}\right]$$

Even though Chernoff's bounds are never as good as the best moment bound, in many cases they are easier to handle.

18.2.4 Chebyshev-Cantelli Inequality

Theorem 1: Let $t \geq 0$. Then,

$$P\left(X - \mathbb{E}\left[X\right] \geq t\right) \;\leq\; \frac{Var\left(X\right)}{Var\left(X\right) + t^2}$$

Proof: We may assume without loss of generality that $\mathbb{E}\left[X\right] = 0$. Then for a constant $t \in \mathbb{R}^+$:

$$t \;=\; \mathbb{E}\left[t - X\right] \;\leq\; \mathbb{E}\left[(t - X)\, \mathbb{1}_{\{X<t\}}\right]$$

(Where $\mathbb{1}$ denotes the indicator function). Using Cauchy-Schwarz inequality, $\mathbb{E}\left[YZ\right] \;\leq\; \sqrt{\mathbb{E}\left[Y^2\right]\mathbb{E}\left[Z^2\right]}$ where X and Y are random variables such that $Y \;=\; (t - X)$ and $Z \;=\; \mathbb{1}_{\{X<t\}}$

$$t \;\leq\; \mathbb{E}\left[(t - X)\, \mathbb{1}_{\{X<t\}}\right] \;\leq\; \sqrt{\mathbb{E}\left[(t - X)^2\right]\mathbb{E}\left[\mathbb{1}^2_{\{X<t\}}\right]}$$

$$\Rightarrow\; t^2 \;\leq\; \mathbb{E}\left[(t - X)^2\right]\mathbb{E}\left[\mathbb{1}^2_{\{X<t\}}\right]$$

$$\Rightarrow\; t^2 \;\leq\; \mathbb{E}\left[(t - X)^2\right] P\left(X < t\right)$$

$$\Rightarrow\; t^2 \;\leq\; \left(Var\left\{X\right\} + t^2\right) P\left(X < t\right)$$

$$\Rightarrow\; P\left(X < t\right) \;\geq\; \frac{t^2}{Var\left\{X\right\} + t^2}$$

$$\Rightarrow\; P\left(X > t\right) \;\geq\; 1 - \frac{t^2}{Var\left\{X\right\} + t^2}$$

$$\Rightarrow\; P\left(X > t\right) \;\geq\; \frac{Var\left\{X\right\}}{Var\left\{X\right\} + t^2}$$

18.2.5 Chebyshev-association Inequality

Theorem 2: *Let f and g be nondecreasing real-valued functions defined on the real line. If X is a real-valued random variable, then*

$$\mathbb{E}\left[f(X)g(X)\right] \;\geq\; \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(X)\right]$$

Proof: Let Y be another random variable with same distribution as X but independent of it.

$$\Rightarrow\; \left(f\left(X\right) - f\left(Y\right)\right)\left(g\left(X\right) - g\left(Y\right)\right) \;\geq\; 0$$

$$\Rightarrow\; \mathbb{E}\left[\left(f\left(X\right) - f\left(Y\right)\right)\left(g\left(X\right) - g\left(Y\right)\right)\right] \;\geq\; 0$$

$$\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right) + f\left(Y\right)g\left(Y\right) - f\left(X\right)g\left(Y\right) - f\left(Y\right)g\left(X\right)\right] \;\geq\; 0$$

$$\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right)\right] + \mathbb{E}\left[f\left(Y\right)g\left(Y\right)\right] \;\geq\; \mathbb{E}\left[f\left(X\right)g\left(Y\right)\right] + \mathbb{E}\left[f\left(Y\right)g\left(X\right)\right]$$

$$\Rightarrow\; 2\,\mathbb{E}\left[f\left(X\right)g\left(X\right)\right] \;\geq\; 2\,\mathbb{E}\left[f\left(X\right)\right]\mathbb{E}\left[g\left(X\right)\right]$$

(as X and Y are independent of each other)

$$\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right)\right] \;\geq\; \mathbb{E}\left[f\left(X\right)\right]\mathbb{E}\left[g\left(X\right)\right]$$

Theorem 3: *Let f be non-decreasing and g be non-increasing real-valued functions defined on the real line. If X is a real-valued random variable, then*

$$\mathbb{E}\left[f(X)g(X)\right] \;\leq\; \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(X)\right]$$

Proof: Let Y be another random variable with same distribution as X but independent of it.

$$\begin{aligned}
\Rightarrow\; \left(f\left(X\right)-f\left(Y\right)\right)\left(g\left(X\right)-g\left(Y\right)\right) \;&\leq\; 0\\
\Rightarrow\; \mathbb{E}\left[\left(f\left(X\right)-f\left(Y\right)\right)\left(g\left(X\right)-g\left(Y\right)\right)\right] \;&\leq\; 0\\
\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right)+f\left(Y\right)g\left(Y\right)-f\left(X\right)g\left(Y\right)-f\left(Y\right)g\left(X\right)\right] \;&\leq\; 0\\
\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right)\right]+\mathbb{E}\left[f\left(Y\right)g\left(Y\right)\right] \;&\leq\; \mathbb{E}\left[f\left(X\right)g\left(Y\right)\right]+\mathbb{E}\left[f\left(Y\right)g\left(X\right)\right]\\
\Rightarrow\; 2\,\mathbb{E}\left[f\left(X\right)g\left(X\right)\right] \;&\leq\; 2\,\mathbb{E}\left[f\left(X\right)\right]\mathbb{E}\left[g\left(X\right)\right]\\
&\qquad\text{(as } X \text{ and } Y \text{ are independent of each other)}\\
\Rightarrow\; \mathbb{E}\left[f\left(X\right)g\left(X\right)\right] \;&\leq\; \mathbb{E}\left[f\left(X\right)\right]\mathbb{E}\left[g\left(X\right)\right]
\end{aligned}$$

18.3 Sums of independent random variables

In this introductory section we recall some simple inequalities for sums of independent random variables. Here we are primarily concerned with upper bounds for the probabilities of deviations from the mean, that is, to obtain inequalities for $P\left\{S_n - \mathbb{E}\,S_n \geq t\right\}$, with $S_n = \sum_{i=1}^{n} X_i$, where $X_1, X_2, ..., X_n$ are independent real-valued random variables. Chebyshev's inequality and independence immediately imply

$$P\left\{S_n - \mathbb{E}\,S_n \geq t\right\} \;\leq\; \frac{Var\left\{S_n\right\}}{t^2} \;=\; \frac{\sum_{i=1}^{n} Var\left\{X_i\right\}}{t^2}$$

$$\Rightarrow\; P\left\{\frac{S_n}{n} - \mathbb{E}\,\frac{S_n}{n} \geq t\right\} \;\leq\; \frac{Var\left\{\frac{S_n}{n}\right\}}{t^2} \;=\; \frac{\sum_{i=1}^{n} Var\left\{X_i\right\}}{n^2 t^2}$$

Writing $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} Var\left\{X_i\right\}$ (when σ_i is same for all random variables X_i),

$$\Rightarrow\; P\left\{\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\,X_i \geq t\right\} \;\leq\; \frac{\sigma^2}{nt^2}$$

This simple inequality is at the basis of the weak law of large numbers. To understand why this inequality is unsatisfactory, recall that, under some additional regularity conditions, the central limit theorem states that

18.4 References:

"Concentration-of-measure inequalities", Lecture notes by Gabor Lugosi, June 25, 2009