CPSC 536N: Randomized Algorithms 2011-12 Term 2

Lecture 17

Prof. Nick Harvey *University of British Columbia*

So far we have seen two concentration bounds for scalar random variables: Markov and Chernoff. For our sort of applications, these are by far the most useful. In most introductory probability courses, you are likely to see another inequality, which is **Chebyshev's inequality**. Its strength lies between the Markov and Chernoff inequalities: the concentration bounds we get from Chebyshev is usually better than Markov but worse than Chernoff. On the other hand, Chebyshev requires stronger hypotheses than Markov but weaker hypotheses than Chernoff.

1 Variance

We begin by reviewing **variance**, and other related notions, which should be familiar from an introductory probability course. The variance of a random variable X is

$$\mathrm{Var}[X] \;=\; \mathrm{E}\Big[\big(X-\mathrm{E}[X]\big)^2\Big] \;=\; \mathrm{E}[X^2]-\mathrm{E}[X]^2.$$

The **covariance** between two random variables X and Y is

$$\mathrm{Cov}[X,Y] \;=\; \mathrm{E}\Big[\big(X-\mathrm{E}[X]\big)\big(Y-\mathrm{E}[Y]\big)\Big] \;=\; \mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y].$$

This gives some measure of the correlation between X and Y.

Here are some properties of variance and covariance that follow from the definitions by simple calculations.

Claim 1 *If X and Y are independent then* $\mathrm{Cov}[X,Y]=0$.

Claim 2 $\mathrm{Cov}[X+Y,Z]=\mathrm{Cov}[X,Z]+\mathrm{Cov}[Y,Z]$.

More generally, induction shows

Claim 3 $\mathrm{Cov}[\sum_i X_i,Z]=\sum_i \mathrm{Cov}[X_i,Z]$.

Claim 4 $\mathrm{Var}[X+Y]=\mathrm{Var}[X]+\mathrm{Var}[Y]+2\cdot\mathrm{Cov}[X,Y]$.

More generally, induction shows

Claim 5 *Let $X_1,\ldots,X_n$ be arbitrary random variables. Then*

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \;=\; \sum_{i=1}^n \mathrm{Var}[X_i]+2\sum_{i=1}^n\sum_{j>i}\mathrm{Cov}[X_i,X_j].$$

In particular,

Claim 6 *Let $X_1,\ldots,X_n$ be mutually independent random variables. Then* $\mathrm{Var}[\sum_{i=1}^n X_i] \;=\sum_{i=1}^n \mathrm{Var}[X_i]$.

2 Chebyshev's Inequality

Chebyshev's inequality you've also presumably seen before. It is a 1-line consequence of Markov's inequality.

Theorem 7 *For any $t > 0$,*

$$\Pr\Big[\big|X - \mathrm{E}[X]\big| \geq t\Big] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}.$$

PROOF:

$$\Pr\Big[\big|X - \mathrm{E}[X]\big| \geq t\Big] \;=\; \Pr\Big[\big(X - \mathrm{E}[X]\big)^2 \geq t^2\Big] \;\leq\; \frac{\mathrm{E}[\big(X - \mathrm{E}[X]\big)^2]}{t^2} \;=\; \frac{\mathrm{Var}[X]}{t^2},$$

where the inequality is by Markov's inequality. □

As a quick example, suppose we independently flip n fair coins. What's the probability that we see at least $3n/4$ heads? Let X_i be the indicator random variable of the event "ith toss is heads". Let $\mu = \mathrm{E}[\sum_i X_i] = n/2$. So we want to analyze $\Pr[\sum_i X_i - \mu \geq n/4]$.

Bound from Chebyshev: Note that

$$\mathrm{Var}[X_i] \;=\; \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 \;=\; 1/2 - 1/4 \;=\; 1/4.$$

By independence,

$$\mathrm{Var}[\textstyle\sum_i X_i] \;=\; \displaystyle\sum_i \mathrm{Var}[X_i] \;=\; n/4.$$

By Chebyshev's inequality

$$\Pr[\textstyle\sum_i X_i - \mu \geq n/4] \;\leq\; \Pr\Big[\big|\sum_i X_i - \mathrm{E}[\sum_i X_i]\big| \geq n/4\Big] \;\leq\; \displaystyle\frac{\mathrm{Var}[\sum_i X_i]}{(n/4)^2} \;=\; \frac{n/4}{(n/4)^2} \;=\; \frac{4}{n}.$$

Bound from Chernoff: Chernoff's inequality gives

$$\Pr[\textstyle\sum_i X_i - \mu \geq n/4] \;=\; \Pr[\sum_i X_i - \mu \geq \mu/2] \;\leq\; \exp(-(1/2)^2\mu/3) \;<\; 0.96^n.$$

This is better than the bound from Chebyshev for $n \geq 71$.

So Chebyshev is weaker than Chernoff, at least for analyzing sums of independent Bernoulli trials. So why do we bother studying Chebyshev? One reason is that Chernoff is designed for analyzing sums of **mutually independent** random variables. That is quite a strong assumption. In some scenarios, our random variables are not mutually independent, or perhaps we deliberately choose them not to be mutually independent.

- For example, generating mutually independent random variables requires a lot of random bits and, as discussed last time, randomness is a "precious resource". We will see that decreasing the number of random bits give another method to derandomize algorithms.
- Another important example is in constructing hash functions, which are random-like functions. Generating a completely random function takes a huge number of random bits. So instead we often try to use hash functions involving less randomness.

3 k-wise independence

A set of events $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are called **k-wise independent** if for any set $I \subseteq \{1, \ldots, n\}$ with $|I| \leq k$ we have

$$\Pr[\wedge_{i \in I} \mathcal{E}_i] \; = \; \prod_{i \in I} \Pr[\mathcal{E}_i].$$

The term **pairwise independence** is a synonym for 2-wise independence.

Similarly, a set of discrete random variables $X_1, \ldots, X_n$ are called **k-wise independent** if for any set $I \subseteq \{1, \ldots, n\}$ with $|I| \leq k$ and any values x_i we have

$$\Pr[\; \wedge_{i \in I} \; X_i = x_i \;] = \prod_{i \in I} \Pr[X_i = x_i].$$

Claim 8 *Suppose $X_1, \ldots, X_n$ are **k-wise independent**. Then*

$$\mathrm{E}[\; \textstyle\prod_{i \in I} X_i \;] \; = \; \displaystyle\prod_{i \in I} \mathrm{E}[X_i] \qquad \forall I \text{ with } |I| \leq k.$$

Proof: For notational simplicity, consider the case $I = \{1, \ldots, k\}$. Then

$$\begin{aligned}
\mathrm{E}[\; \textstyle\prod_{i=1}^k X_i \;] \; &= \; \displaystyle\sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \; \Pr[\; \wedge_{i=1}^k \; X_i = x_i \;] \cdot \prod_{i=1}^k x_i \\
&= \; \sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \; \prod_{i=1}^k \Pr[X_i = x_i] \cdot x_i \qquad (k\text{−wise independence}) \\
&= \; \Big(\sum_{x_1} \Pr[X_1 = x_1] \cdot x_1 \Big) \cdots \Big(\sum_{x_k} \Pr[X_k = x_k] \cdot x_k \Big) \\
&= \; \prod_{i=1}^k \mathrm{E}[X_i].
\end{aligned}$$

□

Example: To get a feel for pairwise independence, consider the following three Bernoulli random variables that are pairwise independent but not mutually independent. There are 4 possible outcomes of these three random variables. Each of these outcomes has probability 1/4.

X_1	X_2	X_3
0	0	0
0	1	1
1	0	1
1	1	0

They are certainly not mutually independent because the event $X_1 = X_2 = X_3 = 1$ has probability 0, whereas $\prod_{i=1}^3 \Pr[X_i = 1] = (1/2)^3$. But, by checking all cases, one can verify that they are pairwise independent.

3.1 Constructing Pairwise Independent RVs

Let $\mathbb{F}$ be a finite field and $q = |\mathbb{F}|$. We will construct RVs $Y_1, \cdots, Y_q$ such that each Y_i is uniform over $\mathbb{F}$ and the Y_i's are pairwise independent. To do so, we need to generate only *two* independent RVs X_1 and X_2 that are uniformly distributed over $\mathbb{F}$. We then define

$$Y_i \; = \; X_1 + i \cdot X_2. \tag{1}$$

Claim 9 *Each Y_i is uniformly distributed on $\mathbb{F}$.*

Proof: For $i = 0$ we have $Y_i = X_1$, which is uniform. For $i \neq 0$ and any $j \in \mathbb{F}$ we have

$$\begin{aligned}
\Pr[Y_i = j] &= \Pr[X_1 + i \cdot X_2 = j] \\
&= \Pr[X_2 = (j - X_1)/i] \\
&= \sum_{x \in \mathbb{F}} \Pr[\; X_2 = (j - x)/i \;\wedge\; X_1 = x \;] \qquad \text{(total probability law)} \\
&= \sum_{x \in \mathbb{F}} \Pr[X_2 = (j - x)/i] \cdot \Pr[X_1 = x] \qquad \text{(pairwise independence)} \\
&= (1/q) \cdot \sum_{x \in \mathbb{F}} \Pr[X_2 = (j - x)/i] \\
&= (1/q),
\end{aligned}$$

since as x ranges through $\mathbb{F}$, $(j-x)/i$ also ranges through all of $\mathbb{F}$. (In other words, the map $x \mapsto (j-x)/i$ is a bijection of $\mathbb{F}$ to itself.) So Y_i is uniform. □

Claim 10 *The Y_i's are pairwise independent.*

Proof: We wish to show that, for any distinct RVs Y_i and Y_j and any values $a, b \in \mathbb{F}$, we have

$$\Pr[\; Y_i = a \;\wedge\; Y_j = b \;] \;=\; \Pr[Y_i = a] \cdot \Pr[Y_j = b] \;=\; 1/q^2.$$

This event is equivalent to $X_1 + i \cdot X_2 = a$ and $X_1 + j \cdot X_2 = b$. We can also rewrite that as:

$$\begin{pmatrix} 1 & i \\ 1 & j \end{pmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \;=\; \begin{pmatrix} a \\ b \end{pmatrix}.$$

This holds precisely when

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \;=\; \begin{pmatrix} 1 & i \\ 1 & j \end{pmatrix}^{-1} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \;=\; \begin{pmatrix} ja - ib \\ b - a \end{pmatrix} / (j - i).$$

Since X_1 and X_2 are independent and uniform over $\mathbb{F}$, this event holds with probability $1/q^2$. □

Corollary 11 *Given $2n$ mutually independent, uniformly random bits, we can construct 2^n pairwise independent, uniformly random strings in $\{0,1\}^n$.*

Proof: Apply the previous construction to the finite field $\mathbb{F}_{2^n}$. The $2n$ mutually independent random bits are used to construct X_1 and X_2. The random strings $Y_1, \ldots, Y_{2^n}$ are constructed as in (1). □

3.2 Example: Max Cut with pairwise independent RVs

Once again let's consider the Max Cut problem. We are given a graph $G = (V, E)$ where $V = \{1, \ldots, n\}$. We will choose $\{0,1\}$-valued random variables $Y_1, \ldots, Y_n$. If $Y_i = 1$ then we add vertex i to U.

Our original algorithm chose $Y_1, \ldots, Y_n$ to be mutually independent and uniform. Instead we will pick $Y_1, \ldots, Y_n$ to be *pairwise independent* and uniform. Then

$$\begin{aligned}
\mathrm{E}[|\delta(U)|] &= \sum_{ij \in E} \Pr[\ (i \in U \wedge j \notin U)\ \vee\ (i \notin U \wedge j \in U)\ ] \\
&= \sum_{ij \in E} \Pr[\ i \in U \wedge j \notin U\ ] + \Pr[\ i \notin U \wedge j \in U\ ] \\
&= \sum_{ij \in E} \Pr[Y_i]\Pr[\overline{Y_j}] + \Pr[\overline{Y_i}]\Pr[Y_j] \qquad \text{(pairwise independence)} \\
&= \sum_{ij \in E} \Big((1/2)^2 + (1/2)^2 \Big) \\
&= |E|/2
\end{aligned}$$

So the original algorithm works just as well if we make pairwise independent decisions instead of mutually independent decisions for placing vertices in U. The following theorem shows the advantage of making pairwise independent decisions.

Theorem 12 *There is a deterministic, polynomial time algorithm to find a cut $\delta(U)$ with $|\delta(U)| \geq |E|/2$.*

PROOF: By Corollary 11, we only need $b = O(\log n)$ mutually independent, uniform random bits $X_1, \ldots, X_b$ in order to generate our pairwise independent, uniform random bits $Y_1, \ldots, Y_n$. We have just argued that these pairwise independent Y_i's will give us

$$\mathrm{E}_{X_1,\ldots,X_b}[|\delta(U)|] \;=\; |E|/2.$$

So there must *exist* some particular bits $(x_1, \ldots, x_b)$ such that fixing $X_i = x_i$ for all i, we get $|\delta(U)| \geq |E|/2$. We can deterministically find such bits $(x_1, \ldots, x_b)$ by exhaustive search in $2^b = 2^{O(\log n)} = \mathrm{poly}(n)$ trials. This gives a deterministic, polynomial time algorithm. □

4 Chebyshev with pairwise independent RVs

One of the main benefits of pairwise independent RVs is that Chebyshev's inequality still works beautifully. Suppose that $X_1, \ldots, X_n$ are pairwise independent. For any $i \neq j$,

$$\mathrm{Cov}[X_i, X_j] \;=\; \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j] \;=\; 0,$$

by Claim 8. So

$$\mathrm{Var}[\textstyle\sum_{i=1}^n X_i] \;=\; \displaystyle\sum_{i=1}^n \mathrm{Var}[X_i],$$

by Claim 5. So

$$\Pr[|\textstyle\sum_i X_i - \mathrm{E}[\sum_i X_i]| > t] \;\leq\; \dfrac{\mathrm{Var}[\sum_i X_i]}{t^2}.$$

This is exactly the same bound that we would get if the X_i's were mutually independent.