CS 395T: Sublinear Algorithms Fall 2014

Lecture 2 — September 2, 2014

Prof. Eric Price *Scribe: Ben Leedom*

1 Markov and Chebyshev Inequalities

Let $X \geq 0$ be a nonnegative random variable. For all $t \geq 0$, $\mathbb{E}[X] \geq t \cdot \mathbb{P}[X \geq t]$, so

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[x]}{t}. \tag{1}$$

Although it is trivial, this bound appears all over the place. Hence it gets a name: *Markov's inequality.*

Let $\mu := \mathbb{E}[X], \mathrm{Var}(X) := \mathbb{E}[(x-\mu)^2]$. We have that

$$\mathbb{P}[|x-\mu| \geq t] = \mathbb{P}[(x-\mu)^2 \geq t^2] \leq \frac{\mathrm{Var}(X)}{t^2} \tag{2}$$

which is known as *Chebyshev's inequality.* This will appear very often throughout the course.

In general, one can take arbitrary moments:

$$\mathbb{P}[|x-\mu| \geq t] = \mathbb{P}[|x-\mu|^k \geq t^k] \leq \frac{\mathbb{E}[|x-\mu|^k]}{t^k} \tag{3}$$

and doing so for $k \geq 3$ is known as a *higher moment method.*

As an example for how these moment methods work, consider $X \sim N(0,1)$ (with PDF $p(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$). Then $\mathbb{E}[x^k] \approx k^{\frac{k}{2}}$, so the kth moment method gives:

$$\mathbb{P}[x \geq t] \lesssim (\frac{k}{t^2})^{\frac{k}{2}} \tag{4}$$

Therefore the best moment k in this case depends on t: choosing $k = t^2/2$ gives a bound of $\mathbb{P}[x \geq t] \lesssim 2^{-t^2/4}$, while substantially larger or substantially lower values of k give worse bounds.

In most cases, a simpler method than optimizing the choice of k in a higher moment method is to use a *moment generating function.*

2 Moment Generating Functions

Define the moment generating function $\Phi(\lambda)$ like this:

$$\Phi(\lambda) = \mathbb{E}[e^{\lambda(x-\mu)}]$$

By Markov's inequality,

$$\mathbb{P}[x - \mu \geq t] = \mathbb{P}[e^{\lambda(x-\mu)} \geq e^{\lambda t}] \leq \frac{\Phi(\lambda)}{e^{\lambda t}} \tag{5}$$

where we can optimize over λ.

Note that $e^{\lambda x} = 1 + \lambda x + \frac{\lambda^2 x^2}{2} + \frac{\lambda^3 x^3}{3!} + \cdots$ is a weighted average of the moments of x. Therefore optimizing λ in (5) is analogous to optimizing k in (4). It's actually slightly weaker: the best higher moment bound is at least as good as the best moment generating function bound. However, we shall see that the moment generating bound is much more convenient, and usually does a good enough job.

Example: Gaussian Variables. Consider the normal distribution $N(0, \sigma^2)$, given by the PDF $p(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}$. The moment generating function is

$$\Phi(\lambda) = \int_{-\infty}^{-\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} e^{\lambda t} dt = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[\frac{(t-\sigma^2\lambda)^2}{2\sigma^2} + \frac{\sigma^2\lambda^2}{2}]} dt = e^{\frac{\sigma^2\lambda^2}{2}} \tag{6}$$

Thus

$$\mathbb{P}[x \geq t] \leq \inf_{\lambda} \frac{\Phi(\lambda)}{e^{\lambda t}} = \inf_{\lambda} e^{\frac{\lambda^2\sigma^2}{2} - \lambda t} = \inf_{\lambda} e^{\frac{1}{2}(\lambda\sigma - \frac{t}{\sigma})^2} \cdot e^{-\frac{t^2}{2\sigma^2}}$$

But of course $\inf_\lambda e^{\frac{1}{2}(\lambda\sigma - \frac{t}{\sigma})^2} = 1$, so this gives us

$$\mathbb{P}[x \geq t] \leq e^{-\frac{t^2}{2\sigma^2}} \tag{7}$$

with the choice of $\lambda = \frac{t}{\sigma^2}$.

3 Subgaussian variables

We say a random variable is "subgaussian" if its moment generating function is bounded by that of a gaussian:

Definition 1. *A variable X is* subgaussian with parameter σ *if for all λ,*

$$\Phi(\lambda) \leq e^{\frac{\lambda^2\sigma^2}{2}}.$$

As with (7), for subgaussian X we have

$$\mathbb{P}[x \geq \mu + t] \leq e^{-\frac{t^2}{2\sigma^2}}$$

and (by replacing X with $-X$) we also have

$$\mathbb{P}[x \leq \mu - t] \leq e^{-\frac{t^2}{2\sigma^2}}.$$

Example: bounded random variables. If $x \in \{-1, 1\}$ uniformly at random, then x is subgaussian with parameter $\sigma = 1$. More generally, if $x \in [a, b]$, then x is subgaussian with parameter $\sigma = \frac{b-a}{2}$. The proof is left as an exercise.

If X_1 and X_2 are independent and subgaussian with parameters σ_1 and σ_2, then $X_1 + X_2$ is subgaussian with parameter $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Proof:

$$\mathbb{E}[e^{\lambda(X_1+X_2)}] = \mathbb{E}[e^{\lambda X_1} e^{\lambda X_2}] = \mathbb{E}[e^{\lambda X_1}]\,\mathbb{E}[e^{\lambda X_2}] \leq e^{\lambda^2\sigma_1^2} e^{\lambda^2\sigma_2^2} = e^{\lambda^2(\sigma_1^2+\sigma_2^2)}$$

This property is why the moment generating function is so nice to work with.

Example: coin flips. suppose we flip n coins, getting $x_1, \ldots, x_n \in \{0, 1\}$. We expect the number of heads $\sum x_i$ to be about $n/2$, but how well does it concentrate about the mean? What is a bound on

$$\mathbb{P}[\Sigma \geq (\frac{1}{2} + \epsilon)n]?$$

Well, each x_i is bounded, so it is subgaussian with parameter $\sigma = \frac{1}{2}$. Since the x_i are independent, the sum has parameter $\sigma = \frac{\sqrt{n}}{2}$. Therefore

$$\mathbb{P}[\sum x_i \geq \frac{n}{2} + \epsilon n] \leq e^{\frac{-(\epsilon n)^2}{2(\frac{\sqrt{n}}{2})}} = e^{-2\epsilon^2 n}$$

How many samples do we need to determine μ to $\pm\epsilon$ with $(1 - \delta)$ probability?

$$n = \frac{1}{2\epsilon^2} \log \frac{2}{\delta} = \Theta(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$$

The logic in this example is general, giving the *hoeffding bound*:

Theorem 2 (Hoeffding)**.** *Let X_i for $i = 1, \ldots, n$ be independent subgaussian variables with mean μ_i and parameter σ_i (for example, be bounded in the range $[a_i - \sigma_i, a_i + \sigma_i]$). Then*

$$\mathbb{P}[\sum_i (x_i - \mu_i) > t] \leq e^{-\frac{t^2}{2\sum_i \sigma_i^2}}$$

4 Alternate definitions of subgaussian variables

One might try to define "subgaussian" variables to be ones that are bounded by a Gaussian in other ways than the moment generating function. For instance, you might consider bounding the tail probability or the individual moments:

(Bounded Tail): X is subgaussian if for some constant C and parameter σ, with $Z \sim N(0, \sigma^2)$ and for all t,

$$\mathbb{P}[|X| \geq t] \leq C\,\mathbb{P}[|Z| \geq t].$$

(Moments): X is subgaussian if for some parameter θ, for all k we have

$$\mathbb{E}[X^{2k}] \leq \frac{(2k)!}{(2^k k!)}\theta^{2k}.$$

Fortunately, these definitions are *equivalent* (up to constants). So it suffices to prove that any of these properties hold.

5 Sub-exponential variables

Some variables do not concentrate as well as a gaussian. A useful subset of these are variables that concentrate as well as *exponential* random variables.

Definition 3. *A variable X is* sub-exponential with parameters σ, B *if* $\mathbb{E}[e^{\lambda(x-\mu)}] \leq e^{\frac{\sigma^2\lambda^2}{2}}$ *for all* $|\lambda| \leq B$.

Example: exponential random variable. Let X be a continuous variable with $p(x) = ce^{-cx}$, for $x \geq 0$, so that $\mathbb{E}[x] = \frac{1}{c}$. Then

$$\mathbb{E}[e^{\lambda x}] = c\int_0^\infty e^{\lambda x}e^{-cx}dx = \frac{c}{c-\lambda}$$

when $\lambda < c$, and so

$$\mathbb{E}[e^{\lambda(x-\mu)}] = \frac{ce^{-\frac{\lambda}{c}}}{c-\lambda}$$

Plotting it, we see that for all $|\lambda| \leq \frac{1}{2}$.

$$\mathbb{E}[e^{\lambda(x-\mu)}] \leq e^{\lambda^2}.$$

Therefore X is subexponential with parameters $(\sqrt{2}, 1/2)$.

Example: square of a gaussian. Example: let $Z \sim N(0,1)$ and $X = Z^2$.

$$\mathbb{E}[e^{\lambda(x-1)}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\infty} e^{\lambda(z^2-1)}e^{-\frac{z^2}{2}}dz = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \tag{8}$$

Plotting the function, we can observe that for all $|\lambda| \leq \frac{1}{4}$,

$$\mathbb{E}[e^{\lambda(x-1)}] \leq e^{\frac{4\lambda^2}{2}}.$$

Therefore X is subexponential with parameters $(2, 1/4)$.

Using the same idea as the proof of (7), but requiring $|\lambda| \in [-B, B]$, we get a tail bound that contains two terms: the same as for subgaussian random variables if the optimal $\lambda = t/\sigma^2 \in$

$[-B, B]$, and a simple exponential otherwise. In particular, we have

$$\mathbb{P}[x \geq \mu + t] \leq \begin{cases} e^{-\frac{t^2}{2\sigma^2}} & \text{if } 0 \leq t \leq B\sigma^2 \\ e^{-\frac{Bt}{2}} & \text{if } t \geq B\sigma^2 \end{cases}$$

and a similar bound for the lower tail, after negating t.

Just as for subgaussians, if X_1, X_2 are sub-exponential with parameters $(\sigma_1, B_1), (\sigma_2, B_2)$, then $X_1 + X_2$ is sub-exponential with parameters $(\sqrt{\sigma_1^2 + \sigma_2^2}, \min(B_1, B_2))$.

6 Johnson-Lindenstrauss Theorem

Lemma 4 (Johnson-Lindenstrauss '84)**.** *Let $x_1, \ldots, x_n \in \mathbb{R}^d$. There exist $y_1, \ldots, y_n \in \mathbb{R}^m$ such that*

$$\|y_i - y_j\|_2 = (1 \pm \epsilon)\|x_i - x_j\|_2 \forall i, j$$

with $m = O(\frac{1}{\epsilon^2} \log n)$. Moreover, such y_i can be computed efficiently.

Proof. Let $A \in \mathbb{R}^{m \times d}$, $A_{ij} \sim N(0, \frac{1}{m})$. We simply set $y_i = Ax_i$.

To see this works, consider any $z \in \mathbb{R}^d$. We have for each coordinate i that

$$(Az)_i = \sum_j A_{ij} z_i \sim N(0, \frac{\|z\|_2^2}{m}),$$

and the coordinates are independent.

Let $w = N(0, I_m) = (Az) \cdot \frac{\sqrt{m}}{\|z\|_2}$. We would like to show that $\|w\|_2^2$ concentrates about its mean.

Recall from the example that for each i, w_i^2 is the square of a standard normal, and hence subexponential with parameters $(2, 1/4)$. Therefore $\|w\|_2^2 = \sum w_i^2$ is sub-exponential with parameters $(2\sqrt{m}, \frac{1}{4})$. Then

$$\mathbb{P}[\|w\|_2^2 \geq m + \epsilon m] \leq e^{\frac{-\epsilon^2 m^2}{2 \cdot 4m}} = e^{\frac{-\epsilon^2 m}{8}}$$

So $\|Az\|_2^2 = (1 \pm \epsilon)\|z\|_2^2$ with probability $1 - 2e^{\frac{-\epsilon^2 m}{8}}$. This is true for any z, including $z = x_i - x_j$ for each i, j. Setting $m = \frac{8}{\epsilon^2} \log \frac{n^3}{2} = \Theta(\frac{1}{\epsilon^2} \log n)$ gives probability $1 - \frac{1}{n}$ that for all i, j

$$\|y_i - y_j\|_2^2 = (1 \pm \epsilon)\|x_i - x_j\|_2^2$$

□