Sum of Random Variables (and Other "Uncontrolled" Things)

If $X_1, X_2, \ldots, X_n$ are i.i.d. then

$$ Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} $$

(Convergence in distrib.)

means that $P(Z_n \to Z) \to P(N(0,1)) \Rightarrow P(Z = R)$. 

Conjecture: quadratic version of [Berry-Esseen].

Today get useful quadratic bounds that are useful for algorithm design.

Basic Inequalities:

- [Markov]: Any nonnegative r.v. $X$ satisfies $P(X \geq \lambda) \leq \frac{E[X]}{\lambda}$.
- [Chebyshev]: $X$ has mean $\mu$, variance $\sigma^2 = E[(X - \mu)^2] = E[X^2] - E[X]^2$. Then $P(|X - \mu| > \lambda) \leq \frac{\sigma^2}{\lambda^2}$.

If $X$ was itself the sum $X_1 + X_2 + \ldots + X_n$ which were i.i.d. (say) under the same distribution $\text{Ber}(\frac{1}{2})$.

Then $\mu = n \cdot \frac{1}{2}$, $\sigma^2 = \frac{n}{4}$, \text{Var. of exp.}

$$ P(X \geq \frac{n}{2} + t\sqrt{n}) \leq \frac{n}{n t + \frac{1}{4}} \text{ by Markov.} \leq \frac{1}{4t^2} \text{ by Chebyshev.} $$

Note that Chebyshev just requires pairwise independence between r.v.s.

Similar to the analysis of Chebyshev, what about higher independence?

$$ P(X \geq \lambda) = P\left[(X - \mu)^4 \geq \lambda^4 \right] \leq \frac{E[(X - \mu)^4]}{\lambda^4} $$

$$ = \sum_{i \neq j} E[X_i - \mu]^4 + 3 \frac{\sigma^4}{n} + \sum_{i \neq j} \sum_{k \neq j} E[(X_i - \mu)^2 (X_j - \mu)^2] + \ldots + c \frac{\sigma^4}{n} \sum_{i \neq j} (X_i - \mu)^2 (X_j - \mu)^2 + \ldots$$

$$ = \sum (X_i - \mu)^4 + 3n(n-1) \sum_{i \neq j} E[X_i - \mu]^2 E[X_j - \mu]^2$$
Theorem: Let $X_1, X_2, \ldots, X_n$ be independent random variables in $[0,1]$, let $X = \sum X_i$.

Let $\mu := E[X]$, and then

\[
P\left(X > \mu + \lambda\right) \leq \exp\left(-\frac{\lambda^2}{2\mu+\lambda}\right)
\]

\[
P\left(X < \mu - \lambda\right) \leq \exp\left(-\frac{\lambda^2}{3\mu}\right).
\]

There are many beta versions, but this one suffices most of the time. We'll talk about improvements soon.

Application: If $X = \text{Bin}(n, \frac{1}{2})$ then $\mu = \frac{n}{2}$. This says that

\[
P\left[X > \frac{n}{2} + t\sqrt{\frac{n}{2}}\right] \leq \exp\left(-\frac{nt^2}{n + t^2}\right) \quad \text{say } t \ll \sqrt{n}.
\]

\[
\leq \exp\left(-\frac{nt^2}{2n}\right) \leq \exp\left(-\frac{t^2}{2}\right).
\]

\[
\Rightarrow \frac{1}{n} \text{ if } t = \sqrt{2}\ln n.
\]

Application: If $n$ loads into $n$ bins, then known: $E[\text{max load}] = O\left(\frac{\ln n}{\ln \ln n}\right)$.

This says: Load on any bin $i = \text{Bin}(n, \frac{1}{n})$. $E[X] = \mu = 1$.

\[
P\left[X > 1 + t\right] \leq \exp\left(-\frac{t^2}{2 + t}\right) \leq \exp\left(-\frac{t}{3}\right)
\]

\[
\leq \frac{1}{t^{2/3}}(\ln n) \text{ if } t = O(\ln n)
\]

\[
\Rightarrow \text{unimodal overall bins,}
\]

\[
P_i[\text{Bin} \geq \Omega(\ln n) \text{ fails}] \leq \frac{1}{n} \text{ with high probability}.
\]

[Nb: a better concentration bound can give you $\frac{\ln^2 n}{\ln \ln n}$ instead.]
Pf: Say want $P[X > \mu + \lambda]$ = $P[X > m]$. 

Also assume: $X_i \in \{0, 1\}$, will extend soon. \[
P[X_i = 1] = p_i \Rightarrow \mu = \sum p_i
\]

\[
P[X > m] = P[e^{tx} > e^{tm}] \quad \text{for } t > 0.
\]
\[
\leq \frac{E[e^{tx}]}{e^{tm}} = e^{-tm} \prod_{i=1}^{n} E[e^{tx_i}]
\]
\[
= e^{-tm} \prod_{i=1}^{n} (1 - p_i + p_i e^t).
\]
\[
= e^{-tm} (1 + e^{\mu(1 - e^t)})^{-n}
\]
\[
\leq \exp \left\{ -tm + \mu(1 - e^{-t}) \right\}.
\]

How to choose $t \in \mathbb{R}_+$ to minimize this?

Differentiating, we get that $e^{(\mu t - t + \mu) [\ln(1 - e^{-t})]'} = 0 \Rightarrow e^t = \frac{\mu}{\ln(1 + x)}$.

\[
\leq \exp \left\{ \frac{\mu \ln(\mu \lambda) + (n - \mu - \lambda)}{\lambda} \right\}
\]

Plug in $\ln(\mu \lambda) = \ln(1 + \frac{\lambda}{\mu})$
\[
= \frac{\ln(1 + x)}{x}
\]
\[
\leq \exp \left\{ -\lambda t(\lambda + \mu) + \lambda \right\}
\]
\[
= \exp \left\{ -\frac{\lambda^2}{2\mu + \lambda} \right\}.
\]

The proof for the "lower tail" is very similar.

Fact: sometimes the break in the box is "much" better.

\[
P_{\lambda}[X > \mu + \lambda] \leq \left( \frac{e^{\lambda}}{(1 + \frac{\lambda}{2\mu})(1 + \mu)} \right) = \left( \frac{e^{\mu}}{(1 + \mu)(1 + \mu)} \right) \quad \text{if } \lambda = \mu \beta
\]
If \( \text{Bin}(n, p) \) then set \( p = \lambda = \frac{\log n}{\log \log n} \) we get \((1 + \lambda)^{1+\lambda} \leq \frac{1}{\log(n)}\). 

\[
P_0[X \geq 1 + \lambda] \leq \frac{1}{\log(n)} \quad \text{which is the right answer}
\]

Not discrete? 
What about \( X_i \in \{0,1\} \) instead of \( X_i \in \{0,1,3\} \). 
Use that since \( E[X_i] = k_i \) then \( E[f(X_i)] \) is maximized when you put \( k_i \) at \( k \) and \( 1 - k \) at 0.

Suppose variables are not independent?

- As long as \( X_i \)'s are negatively correlated (some of the variables "high" makes it more likely for others to be "low").

\[
\Rightarrow \text{these bounds can hold.}
\]

- Formally: \( E[f(X_i : i \in A)] g(X_j : j \in B)] \leq E[f(X_i : i \in A)] E[g(X_j : j \in B)] \)
  for all disjoint \( A, B \), and \( f, g \) are monotone increasing functions.

[Eq: if a bin \( i \) is occupied then bin \( j \neq i \) more likely to be empty.] 
So if \( X_i = \mathbb{I}(i \text{ occupied}) \) then \( X_i \) are negatively correlated [Eq].

\[
\Rightarrow \text{can prove that } \sum X_i = \# \text{occupied bins is a sum of negatively correlated rvs.}
\]

Thm: can then show that \( E[\prod e^{tX_i}] \leq \prod E[e^{tX_i}] \Rightarrow \) rest is proof somehow.

\( \Rightarrow \text{'Chebyshev' bounds hold for these cases too!} \)

Bernstein bounds: \( X_i \in \{0,1\} \) indep, \( \sigma^2 = \sum \sigma_i^2 = \text{Var}(X) \)

\[
P_X[X \geq \mu + \lambda] \leq \exp\left(-\frac{\lambda^2}{2\sigma^2 + \lambda}\right)
\]

\[
P_X[X \leq \mu - \lambda] \leq \exp\left(-\frac{\lambda^2}{2\sigma^2 + \lambda}\right)
\]
BTW: What gives with this \( \exp\left(-\frac{\lambda^2}{2\mu + \beta}\right) \) vs not \( \exp\left(-\frac{\lambda^2}{2\mu}\right) \), etc.

Note: scaling can give no \([-B,B] \) bonds.

Note: Translation takes care of \([-B,B] \) bonds.

Not true! Look at \( n \)-ball auto case. Right side bond gives bond \( \text{Sign} = \text{false} \).

**Beyond Sums:** What about \( f(x_1, x_2, \ldots, x_n) \)? As long as \( f \) is "well-behaved":

1. Want \( x_1, x_2, \ldots, x_n \) to be independent \( \Rightarrow \) product space.
2. Want \( f \) not to depend too much on a single coordinate.

**Drift [Lipschitz function]:** \( f \) is \( C_1 \)-Lipschitz along coordinate \( i \) if

\[
|f(x) - f(x_i)_{\text{flipped to some other value}}| \leq C_i
\]

**Thm:** Let \( f \) be \( C_1 \)-Lipschitz \( \forall i \), \( X \) be product space. Then

\[
\Pr\left[ \exists f \geqEf + \lambda \right] \leq \exp\left(-\frac{\lambda^2}{2\Sigma a_i^2}\right).
\]

Very weak in the case of "balls and bins" \( f \) in the max load in the system.

Essentially depends on the \#4-variables, as opposed to Bernstein-type which are "dimension independent".

**Assume a little more:** get

**Def:** \( f \) is self-bounding if

1. \( f \) is \( C_1 \)-Lipschitz along each coordinate \( \checkmark \) implied (below)
2. \( f \) functions \( f_i \) such that
   
   \[
   (a) \quad 0 \leq f_i(x) - f_i(x_i) \leq 1 \quad \forall x \in \mathbb{R}
   \]
   \[
   (b) \quad \sum_{i=1}^n (f(x) - f_i(x_i)) \leq f(x) \quad \forall x \in \mathbb{R}.
   \]

**Thm:** If \( f \) is self-bounding then

\[
\Pr\left[ \exists f \geq Ef + \lambda \right] \leq \exp\left(-\frac{\lambda^2}{2Ef + \lambda}\right)
\]

\[
\text{[Good example?]}.
\]
Matrix Chernoff: $X_k$ are independent, symmetric matrices of dimension $d$.

Moreover: $X_k \preceq 0$ and $X_k \leq \mathbb{R}I$. \iffalse all $X_k$ are in $[0,1]$ \fi

Let $\mu_{\min} = \lambda_{\min}(\sum E(X_k))$

$\mu_{\max} = \lambda_{\max}(\sum E(X_k))$

Then $P_0 \left[ \lambda_{\max}(\sum X_k) \geq \mu_{\max} + \delta \right] \leq \exp \left( -\frac{\delta^2}{2\mu_{\max} + \delta} \right)$.

Note: if matrices have only values on diagonals, then this is saying that the sum of every diagonal entry is never much more than the max expected diagonal. Obtainable by a union bound. Here we set this for all post matrices.

Also exist Bernstein, Azuma, etc.

Other things to mention: -

- $X$ is a random point in $[0,1]^n$. $\text{Bin}(n, \frac{1}{2})$. Say total UV = 1.

Then $P_0[|X| > \frac{\sqrt{n}}{2} + t\sqrt{n}] = \text{Volume} \setminus \text{Set inside } (\frac{\sqrt{n}}{2} + t\sqrt{n})$

The set $H_\frac{3}{2}$

but $H_1$ is almost 1.

Most measure concentrated around the "equator".

- What about moment bounds? $P_0[X > t] \leq \min_{k \geq 0} \frac{E[X^k]}{t^k}$. 

Phillips & Nelson show that: $\min_{k \geq 0} \frac{E[X^k]}{t^k} \leq \inf_{t > 0} \frac{E[e^{tX}]}{e^{tX}}$ for any non-negative r.v.