Tail Inequalities

Sariel Har-Peled

November 10, 2005
598 - Approximation Algorithms in Geometry

"Wir mssen wissen, wir werden wissen" (We must know, we shall know)
—– David Hilbert

1 Markov Inequality

Theorem 1.1 (Markov Inequality) For a non-negative variable $X$, and $t > 0$, we have:

$$\Pr[X \geq t] \leq \frac{E[X]}{t}.$$

Proof: Assume that this is false, and there exists $t_0 > 0$ such that $\Pr[X \geq t_0] > \frac{E[X]}{t_0}$. However,

$$E[X] = \sum_x x \cdot \Pr[X = x] = \sum_{x < t_0} x \cdot \Pr[X = x] + \sum_{x \geq t_0} x \cdot \Pr[X = x]$$

$$\geq 0 + t_0 \cdot \Pr[X \geq t_0] > 0 + t_0 \cdot \frac{E[X]}{t_0} = E[X],$$

a contradiction. $\blacksquare$

Theorem 1.2 (Chebychev inequality) Let $X$ be a random variable with $\mu_x = E[X]$ and $\sigma_x$ be the standard deviation of $X$. That is $\sigma_x^2 = E[(X - \mu_x)^2]$. Then, $\Pr[|X - \mu_X| \geq t\sigma_X] \leq \frac{1}{t^2}$.

Proof: Note that

$$\Pr[|X - \mu_X| \geq t\sigma_X] = \Pr[(X - \mu_X)^2 \geq t^2\sigma_X^2].$$

Set $Y = (X - \mu_X)^2$. Clearly, $E[Y] = \sigma_X^2$. Now, apply Markov inequality to $Y$. $\blacksquare$

Definition 1.3 Variables $X, Y$ are independent if for any $x, y$ we have:

$$\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y].$$

The following is easy to verify:

Claim 1.4 If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$. If $X$ and $Y$ are independent then $Z = e^X, W = e^Y$ are also independent variables.
2 Tail Inequalities

2.1 The Chernoff Bound — Special Case

**Theorem 2.1** Let $X_1, \ldots, X_n$ be $n$ independent random variables, such that $\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. Then, for any $\Delta > 0$, we have

$$\Pr[Y \geq \Delta] \leq e^{-\Delta^2/2n}.$$ 

**Proof:** Clearly, for an arbitrary $t$, to specified shortly, we have

$$\Pr[Y \geq \Delta] = \Pr[\exp(tY) \geq \exp(t\Delta)] \leq \frac{\mathbb{E}[\exp(tY)]}{\exp(t\Delta)},$$

the first part follows by the fact that $\exp(\cdot)$ preserve ordering, and the second part follows by the Markov inequality.

Observe that

$$\mathbb{E}[\exp(tX_i)] = \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \frac{e^t + e^{-t}}{2} = \frac{1}{2} \left( 1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots \right) + \frac{1}{2} \left( 1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots \right) = \left( 1 + \frac{t^2}{2!} + \cdots + \frac{t^{2k}}{(2k)!} + \cdots \right),$$

by the Taylor expansion of $\exp(\cdot)$. Note, that $(2k)! \geq (k!)2^k$, and thus

$$\mathbb{E}[\exp(tX_i)] = \sum_{i=0}^\infty \frac{t^{2i}}{(2i)!} \leq \sum_{i=0}^\infty \frac{t^{2i}}{2^i(i!)^2} = \sum_{i=0}^\infty \left( \frac{t^2}{2} \right)^i = \exp(t^2/2),$$

again, by the Taylor expansion of $\exp(\cdot)$. Next, by the independence of the $X_i$s, we have

$$\mathbb{E}[\exp(tY)] = \mathbb{E} \left[ \exp \left( \sum_i tX_i \right) \right] = \mathbb{E} \left[ \prod_i \exp(tX_i) \right] = \prod_i \mathbb{E}[\exp(tX_i)] \leq \prod_{i=1}^n e^{t^2/2} = e^{nt^2/2}.$$ 

We have

$$\Pr[Y \geq \Delta] \leq \frac{\exp(nt^2/2)}{\exp(t\Delta)} = \exp(nt^2/2 - t\Delta).$$

Next, by minimizing the above quantity for $t$, we set $t = \Delta/n$. We conclude,

$$\Pr[Y \geq \Delta] \leq \exp \left( \frac{n}{2} \left( \frac{\Delta}{n} \right)^2 - \frac{\Delta}{n} \frac{\Delta}{n} \right) = \exp \left( -\frac{\Delta^2}{2n} \right).$$

By the symmetry of $Y$, we get the following:
Corollary 2.2 Let $X_1, \ldots, X_n$ be $n$ independent random variables, such that $\Pr[X_i = 1] = \Pr[X_i = -1] = \frac{1}{2}$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{n} X_i$. Then, for any $\Delta > 0$, we have

$$\Pr[|Y| \geq \Delta] \leq 2e^{-\Delta^2/2n}.$$ 

Corollary 2.3 Let $X_1, \ldots, X_n$ be $n$ independent coin flips, such that $\Pr[X_i = 0] = \Pr[X_i = 1] = \frac{1}{2}$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{n} X_i$. Then, for any $\Delta > 0$, we have

$$\Pr\left[\left|Y - \frac{n}{2}\right| \geq \Delta\right] \leq 2e^{-2\Delta^2/n}.$$ 

Remark 2.4 Before going any further, it is might be instrumental to understand what this inequalities imply. Consider then case where $X_i$ is either zero or one with probability half. In this case $\mu = \mathbb{E}[Y] = n/2$. Set $\delta = t\sqrt{n}$ ($\sqrt{\mu}$ is approximately the standard deviation of $X$ if $p_i = 1/2$). We have by

$$\Pr\left[\left|Y - \frac{n}{2}\right| \geq \Delta\right] \leq 2 \exp\left(-2\frac{\Delta^2}{n}\right) = 2 \exp\left(-2\frac{(t\sqrt{n})^2}{n}\right) = 2 \exp\left(-2t^2\right).$$

Thus, Chernoff inequality implies exponential decay (i.e., $\leq 2^{-t}$) with $t$ standard deviations, instead of just polynomial (like the Chebychev’s inequality).

2.2 The Chernoff Bound — General Case

Here we present the Chernoff bound in a more general settings.

Question 2.5 Let

1. $X_1, \ldots, X_n$ - $n$ independent Bernoulli trials, where

   \[\Pr[X_i = 1] = p_i, \text{ and } \Pr[X_i = 0] = q_i = 1 - p_i.\]

   Each $X_i$ is known as a Poisson trials.

2. $X = \sum_{i=1}^{b} X_i$. $\mu = \mathbb{E}[X] = \sum_{i} p_i$.

   Question: Probability that $X > (1 + \delta)\mu$?

Theorem 2.6 For any $\delta > 0$, we have $\Pr[X > (1 + \delta)\mu] < \left(\frac{e^{\delta}}{(1 + \delta)^{1+\delta}}\right)^{\mu}$.

Or in a more simplified form, for any $\delta \leq 2e - 1$,

$$\Pr[X > (1 + \delta)\mu] < \exp\left(-\mu\delta^2/4\right),$$

and

$$\Pr[X > (1 + \delta)\mu] < 2^{-\mu(1+\delta)},$$

for $\delta \geq 2e - 1$. 

3
Proof: We have \( \Pr[X > (1 + \delta)\mu] = \Pr[e^{tX} > e^{t(1+\delta)\mu}] \). By the Markov inequality, we have:

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

On the other hand,

\[
\mathbb{E}[e^{tX}] = \mathbb{E}[e^{t(X_1 + X_2 + \ldots + X_n)}] = \mathbb{E}[e^{tX_1}] \cdot \cdots \cdot \mathbb{E}[e^{tX_n}].
\]

Namely,

\[
\Pr[X > (1 + \delta)\mu] < \prod_{i=1}^{n} \frac{\mathbb{E}[e^{tX_i}]}{e^{t(1+\delta)\mu}} = \prod_{i=1}^{n} \frac{(1 - p_i)e^0 + p_i e^t}{e^{t(1+\delta)\mu}} = \prod_{i=1}^{n} \frac{(1 + p_i(e^t - 1))}{e^{t(1+\delta)\mu}}.
\]

Let \( y = p_i(e^t - 1) \). We know that \( 1 + y < e^y \) (since \( y > 0 \)). Thus,

\[
\Pr[X > (1 + \delta)\mu] < \prod_{i=1}^{n} \frac{\exp(p_i(e^t - 1))}{e^{t(1+\delta)\mu}} = \frac{\exp(\sum_{i=1}^{n} p_i(e^t - 1))}{e^{t(1+\delta)\mu}} = \frac{\exp(e^t - 1)}{e^{t(1+\delta)}} = \left( \frac{\exp(\delta)}{(1 + \delta)^{1+\delta}} \right)^\mu,
\]

if we set \( t = \log(1 + \delta) \).

For the proof of the simplified form, see Section 2.3

Definition 2.7 \( F^+(\mu, \delta) = \left[ e^{\delta} (1 + \delta)^{1+\delta} \right]^\mu \).

Example 2.8 Arkansas Aardvarks win a game with probability \( 1/3 \). What is their probability to have a winning season with \( n \) games. By Chernoff inequality, this probability is smaller than

\[
F^+(n/3, 1/2) = \left[ e^{1/2} (1.515)^{n/3} \right] = (0.89745)^{n/3} = 0.964577^n.
\]

For \( n = 40 \), this probability is smaller than 0.236307. For \( n = 100 \) this is less than 0.027145. For \( n = 1000 \), this is smaller than \( 2.17221 \cdot 10^{-16} \) (which is pretty slim and shady). Namely, as the number of experiments is increases, the distribution converges to its expectation, and this converge is exponential.

Theorem 2.9 Under the same assumptions as Theorem 2.6, we have:

\[
\Pr[X < (1 - \delta)\mu] < e^{-\mu \delta^2/2}.
\]

Definition 2.10 \( F^-(\mu, \delta) = e^{-\mu \delta^2/2} \).

\( \Delta^- (\mu, \varepsilon) \) - what should be the value of \( \delta \), so that the probability is smaller than \( \varepsilon \).

\[
\Delta^- (\mu, \varepsilon) = \sqrt{\frac{2 \log 1/\varepsilon}{\mu}}
\]

For large \( \delta \):

\[
\Delta^+ (\mu, \varepsilon) < \frac{\log_2 (1/\varepsilon)}{\mu} - 1
\]
<table>
<thead>
<tr>
<th>Values</th>
<th>Probabilities</th>
<th>Inequality</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>−1, +1</td>
<td>( \Pr[X_i = -1] = )</td>
<td>( \Pr[Y \geq \Delta] \leq e^{-\Delta^2/2n} )</td>
<td>Theorem 2.1</td>
</tr>
<tr>
<td></td>
<td>( \Pr[X_i = 1] = \frac{1}{2} )</td>
<td>( \Pr[Y \leq -\Delta] \leq e^{-\Delta^2/2n} )</td>
<td>Theorem 2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \Pr[</td>
<td>Y</td>
</tr>
<tr>
<td>0, 1</td>
<td>( \Pr[X_i = 0] = )</td>
<td>( \Pr[Y \geq \frac{n}{2}] \geq \Delta] \leq 2e^{-2\Delta^2/n} )</td>
<td>Corollary 2.3</td>
</tr>
<tr>
<td></td>
<td>( \Pr[X_i = 1] = \frac{1}{2} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0, 1</td>
<td>( \Pr[X_i = 0] = 1 - p_i )</td>
<td>( \Pr[Y &gt; (1 + \delta)\mu] &lt; \left( \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^\mu )</td>
<td>Theorem 2.6</td>
</tr>
<tr>
<td></td>
<td>( \Pr[X_i = 1] = p_i )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| For \( \delta \leq 2e - 1 \) | \( \Pr[Y > (1 + \delta)\mu] < \exp(-\mu\delta^2/4) \) | Theorem 2.6 |
| For \( \delta \geq 0 \) | \( \Pr[Y > (1 + \delta)\mu] < 2^{-\mu(1+\delta)} \) |       |
| For \( \delta \geq 0 \) | \( \Pr[Y < (1 - \delta)\mu] < \exp(-\mu\delta^2/2) \) | Theorem 2.9 |

Table 1: Summary of Chernoff type inequalities covered. Here we have \( n \) variables \( X_1, \ldots, X_n \), \( Y = \sum_i X_i \) and \( \mu = \mathbb{E}[Y] \).

2.3 A More Convenient Form

Proof: (of simplified form of Theorem 2.6) Eq. (2) is just Exercise 4.1. As for Eq. (1), we prove this only for \( \delta \leq 1/2 \). For details about the case \( 1/2 \leq \delta \leq 2e - 1 \), see [MR95]. By Theorem 2.6, we have

\[
\Pr[X > (1 + \delta)\mu] < \left( \frac{e^{\delta}}{(1 + \delta)^{1+\delta}} \right)^\mu = \exp(\mu(\delta - \mu(1 + \delta) \ln(1 + \delta))).
\]

The Taylor expansion of \( \ln(1 + \delta) \) is

\[
\delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} - \frac{\delta^4}{4} + \cdots \geq \delta - \frac{\delta^2}{2},
\]

for \( \delta \leq 1 \). Thus,

\[
\Pr[X > (1 + \delta)\mu] < \exp(\mu(\delta - (1 + \delta)(\delta - \frac{\delta^2}{2}))) = \exp(\mu(\delta - \delta + \frac{\delta^2}{2} - \frac{\delta^2}{2} + \delta^3/2)) \leq \exp(\mu(-\frac{\delta^2}{2} + \delta^3/2)) \leq \exp(-\mu\delta^2/4),
\]

for \( \delta \leq 1/2 \).

3 Bibliographical notes

The exposition here follows more or less the exposition in [MR95]. The special symmetric case (Theorem 2.1) is taken from [Cha01], although the proof is only very slightly simpler than the generalized form, it does yield a slightly better constant, and it would be useful when discussing discrepancy.

An orderly treatment of probability is outside the scope of our discussion. The standard text on the topic is the book by Feller [Fel91]. A more accessible text might be any introductory undergrad text on probability, in particular [MN98] has a nice chapter on the topic.

Exercise 4.2 (without the hint) is from [Mat99].
4 Experiments

Exercise 4.1 [2 Points] Prove that for $\delta > 2e - 1$, we have

$$F^+ (\mu, \delta) < \left( \frac{e}{1 + \delta} \right)^{(1+\delta)\mu} \leq 2^{-(1+\delta)\mu}.$$

Exercise 4.2 [10 Points] Let $S = \sum_{i=1}^{n} S_i$ be a sum of $n$ independent random variables each attaining values $+1$ and $-1$ with equal probability. Let $P(n, \Delta) = \Pr[S > \Delta]$. Prove that for $\Delta \leq n/C$,

$$P(n, \Delta) \geq \frac{1}{C} \exp \left( -\frac{\Delta^2}{Cn} \right),$$

where $C$ is a suitable constant. That is, the well-known Chernoff bound $P(n, \Delta) \leq \exp(-\Delta^2/2n)$ is close to the truth.

[Hint: Use Stirling’s formula. There is also an elementary solution, using estimates for the middle binomial coefficients [MN98] pages 83–84], but this solution is considerably more involved and yields unfriendly constants.]

References


