# Lecture Notes: Markov Inequality, Chebyshev Inequality, and Chernoff Bound 

Yufei Tao<br>Chinese University of Hong Kong<br>taoyf@cse.cuhk.edu.hk

10 Feb, 2012

In this lecture, we will study three inequalities that are of paramount importance in analyzing randomized algorithms.

## 1 Markov inequality

Theorem 1 (Markov inequality). Let $X \geq 0$ be a random variable. For any $t>0$, it holds that:

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

Proof. Let $f(X)$ be the probability density function of $X$. It holds that:

$$
\begin{aligned}
\operatorname{Pr}[X \geq t] & =\int_{t}^{\infty} f(X) d X \\
& =\frac{1}{t} \int_{t}^{\infty} t \cdot f(X) d X \\
(\text { as } t>0) & \leq \frac{1}{t} \int_{t}^{\infty} X \cdot f(X) d X \\
(\text { as } X \geq 0) & \leq \frac{1}{t} \cdot \mathbf{E}[X]
\end{aligned}
$$

as needed.
Corollary 1. Let $X \geq 0$ be a random variable. For any $t>0$, it holds that:

$$
\operatorname{Pr}[X \geq t \cdot \mathbf{E}[X]] \leq \frac{1}{t}
$$

The above corollary is perhaps more intuitive (than Theorem 1): it says that the probability for $X$ to be $t$ times larger than its expected value is at most $1 / t$.
Example: sampling. Consider a box of $n$ balls, each of which is either black or white. We want to know the percentage $p$ of the black balls. Each time we can look at a random ball. If we have seen $k$ balls among which $b$ balls are black, we estimate $p$ to be $b / k$. The question is how large $k$ needs to be before our estimate is close to $p$ with a probability close to 1 . To facilitate analysis, let us assume that after drawing a ball, we put it back into the box, so that it may be drawn again with the same chance of any other ball. This is called sampling with replacement.

At the end of the lecture, we will find a fairly good answer to the earlier question, but at this point, we will be content with the following (weaker) claim:

Lemma 1. For any $k \geq 1$, the probability that $b / k \geq 2 p$ is at most $1 / 2$.
Proof. Define random variable $x_{i}(1 \leq i \leq k)$ such that $x_{i}=1$ if the $i$-th ball sampled is black, or 0 otherwise. Thus, $\operatorname{Pr}\left[x_{i}=1\right]=p$, which implies that $\mathbf{E}\left[x_{i}\right]=p$. Clearly, $b=\sum_{i=1}^{k} x_{i}$. Hence:

$$
\mathbf{E}[b]=\mathbf{E}\left[\sum_{i=1}^{k} x_{i}\right]=\sum_{i=1}^{k} \mathbf{E}\left[x_{i}\right]=k p .
$$

By Markov inequality:

$$
\begin{aligned}
\operatorname{Pr}[b \geq 2 \cdot \mathbf{E}[b]] & \leq 1 / 2 \\
\Rightarrow \operatorname{Pr}[b \geq 2 k p] & \leq 1 / 2 \\
\Rightarrow \operatorname{Pr}[b / k \geq 2 p] & \leq 1 / 2 .
\end{aligned}
$$

The lemma indicates that, we can over-estimate $p$ by a factor of 2 with at most $50 \%$ probability, regardless of how many balls are sampled.

## 2 Chebyshev inequality

Theorem 2 (Chebyshev inequality). Let $X$ be a random variable with expectation $\mu$ and variance $\sigma^{2}$ (namely, $\sigma>0$ is the standard deviation). For any $t>0$, it holds that:

$$
\operatorname{Pr}[|X-\mu| \geq t \sigma] \leq \frac{1}{t^{2}}
$$

Proof.

$$
\operatorname{Pr}[|X-\mu| \geq t \sigma]=\operatorname{Pr}\left[(X-\mu)^{2} \geq t^{2} \sigma^{2}\right]
$$

As $\sigma^{2}=\mathbf{E}\left[X^{2}\right]-\mathbf{E}[X]^{2}=\mathbf{E}\left[X^{2}\right]-\mu^{2}$, we have:

$$
\operatorname{Pr}\left[(X-\mu)^{2} \geq t^{2} \sigma^{2}\right]=\operatorname{Pr}\left[(X-\mu)^{2} \geq t^{2}\left(\mathbf{E}\left[X^{2}\right]-\mu^{2}\right)\right]
$$

Define $Y=(X-\mu)^{2}=X^{2}-2 X \mu+\mu^{2}$. Thus:

$$
\begin{aligned}
\mathbf{E}[Y] & =\mathbf{E}\left[X^{2}\right]-2 \mathbf{E}[X] \mu+\mu^{2} \\
& =\mathbf{E}\left[X^{2}\right]-2 \mu^{2}+\mu^{2} \\
& =\mathbf{E}\left[X^{2}\right]-\mu^{2} .
\end{aligned}
$$

Hence:

$$
\operatorname{Pr}\left[(X-\mu)^{2} \geq t^{2}\left(\mathbf{E}\left[X^{2}\right]-\mu^{2}\right)\right]=\operatorname{Pr}\left[Y \geq t^{2} \mathbf{E}[Y]\right]
$$

which is at most $1 / t^{2}$ by Markov inequality.
The theorem says that $X$ can deviate from its expectation by at least $t$ times the standard deviation with probability at most $1 / t^{2}$.

Corollary 2. Let $X$ be a random variable with expectation $\mu$ and variance $\sigma^{2}$ (namely, $\sigma>0$ is the standard deviation). For any $t>0$, it holds that:

$$
\operatorname{Pr}[|X-\mu| \geq t] \leq \frac{\sigma^{2}}{t^{2}}
$$

Sampling (cont.). We now utilize Chebyshev inequality to obtain a stronger claim about the sampling scenario described in the earlier section:

Lemma 2. Let $\delta$ be any value satisfying $0<\delta<1$. With $k=\frac{1}{\epsilon^{2} \delta}$, the probability that $|b / k-p| \leq \epsilon$ is at least $1-\delta$.

Proof. Define random variables $x_{1}, \ldots, x_{k}$ as in the proof of Lemma 1. For each $i \in[1, k], \mathbf{E}\left[x_{i}\right]=p$ and $\operatorname{var}\left[x_{i}\right]=p(1-p)$. As $b=\sum_{i=1}^{k} x_{i}$ and $x_{1}, \ldots, x_{k}$ are mutually independent, we know:

$$
\begin{aligned}
\mathbf{E}[b] & =k p \\
\operatorname{var}[b] & =\sum_{i=1}^{k} \operatorname{var}\left[x_{i}\right]=k p(1-p)
\end{aligned}
$$

Hence:

$$
\begin{aligned}
\operatorname{Pr}[|b / k-p| \geq \epsilon] & =\operatorname{Pr}[|b-k p| \geq \epsilon k] \\
\text { (by Chebyshev inequality) } & \leq \frac{\operatorname{var}[b]}{\epsilon^{2} k^{2}}=\frac{k p(1-p)}{\epsilon^{2} k^{2}}=\frac{p(1-p)}{\epsilon^{2} k} \\
& <1 /\left(\epsilon^{2} k\right) \\
& =\delta
\end{aligned}
$$

Note that the value of $k$ in the above lemma is independent on $n$. In other words, a fixed number of samples is sufficient to achieve the probabilistic guarantee described by the same pair of $\epsilon$ and $\delta$, regardless of the size of population (good news for surveying in a populous country).

## 3 Chernoff bound

Theorem 3 (Chernoff bound). Let $X_{1}, \ldots, X_{k}$ be $k$ independent random variables such that, for each $i \in[1, k]$, $X_{i}$ equals 1 with probability $p$, and 0 with probability $1-p$. Let $X=\sum_{i=1}^{k} X_{i}$ and $\mu=k p$. For any $\epsilon$ satisfying $0<\epsilon<1$, it holds that:

$$
\operatorname{Pr}[|X-\mu| \geq \epsilon \mu] \leq 2 e^{\frac{-\epsilon^{2} \mu}{3}}
$$

We will skip the proof, which can be found in [1] and is a bit technically involved. Remember that this theorem demands $X_{1}, \ldots, X_{k}$ to be independent. When this condition is fulfilled, the Chernoff bound usually gives a tighter bound. Next, we demonstrate this in the sampling scenario of the previous sections.

Sampling (cont.). We will prove the following which significantly improves Lemma 2 when $\delta$ is small.

Theorem 4. Let $\delta$ be any value satisfying $0<\delta<1$. With $k=\frac{3}{\epsilon^{2}} \ln \frac{2}{\delta}$, the probability that $|b / k-p| \leq \epsilon$ is at least $1-\delta$ when $\epsilon<p$.

Proof. Define random variables $x_{1}, \ldots, x_{k}$ as in the proof of Lemma 1 ; recall that they are independent. For each $i \in[1, k], \mathbf{E}\left[x_{i}\right]=p$. As $b=\sum_{i=1}^{k} x_{i}$, we know $\mathbf{E}[b]=k p$. Hence:

$$
\begin{aligned}
\operatorname{Pr}[|b / k-p| \geq \epsilon] & =\operatorname{Pr}[|b-k p| \geq \epsilon k] \\
& =\operatorname{Pr}\left[|b-k p| \geq \frac{\epsilon}{p} k p\right] \\
& =\operatorname{Pr}\left[|b-\mathbf{E}[b]| \geq \frac{\epsilon}{p} \cdot \mathbf{E}[b]\right] \\
\text { (by Chernoff bound) } & =\leq 2 e^{-\frac{\epsilon^{2}}{3 p^{2}} p} \\
& =2 e^{-\frac{\epsilon^{2} k}{3 p}}
\end{aligned}
$$

To make the above at most $\delta$, we need:

$$
k \geq \frac{3 p}{\epsilon^{2}} \ln \frac{2}{\delta}
$$

As $p<1$, taking $k=\frac{3}{\epsilon^{2}} \ln \frac{2}{\delta}$ guarantees the above.

## References

[1] T. Hagerup and C. Rub. A guided tour of chernoff bounds. Information Processing Letters (IPL), 33(6):305-308, 1990.

