Probability: Deviation Inequalities[1]

- **Deviation Inequalities**

  We have seen an example that $\mathbf{Exp}[X]$ may not be anywhere close to what values X can take (recall the $X = 10000$ with 0.5 probability and -10000 with 0.5 probability). Deviation inequalities try to put an *upper bound* on the probability that a random walk deviates too far from the expectation.

  The mother of all deviation inequalities is the following:

  Theorem 1. (Markov's Inequality)

  Let X be a random variable whose range is *non-negative reals*. Then for any $t > 0$, we have

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

  Before we embark on to the proof of Markov's inequality, let us actually understand what it says. For simplicity, assume the probability distribution is uniform (so the expectation is the usual "average"). And also let's fix $t = 2$. Also, just for concreteness, let X denote the height of a random person in a group of people. Then, Markov states that the fraction of people whose height is *at least* twice the average is *at most* $1/2$. Indeed, if not, then more than $1/2$ the fraction will be more than 2 times the average, but that will just drive the average up. The proof below is basically this argument for general probability distributions.

  Proof. By definition of expectation, we have

  $$\mathbf{Exp}[X] = \sum_{k \in \mathbb{R}} k \cdot \mathbf{Pr}[X = k] = \sum_{0 \leq k < t} k \cdot \mathbf{Pr}[X = k] + \sum_{k \geq t} k \cdot \mathbf{Pr}[X = k]$$

  The first summation $\sum_{0 \leq k < t} k \cdot \mathbf{Pr}[X = k] \geq 0$ since all terms are non-negative. The second summation is $\sum_{k \geq t} k \cdot \mathbf{Pr}[X = k] \geq t \cdot \sum_{k \geq t} \mathbf{Pr}[X = k] = t \cdot \mathbf{Pr}[X \geq t]$.

  Putting it all together, we get

  $$\mathbf{Exp}[X] \geq t \cdot \mathbf{Pr}[X \geq t]$$

  which gives what we want by rearrangement. □

  Markov's inequality only talks about non-negative random variables. Indeed, the example in the beginning of this bullet point shows that it cannot be true for general random variables. This is where *variance* comes to play. The following is one of the most general forms of deviation inequalities.

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 28th Aug, 2021*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

Theorem 2. (Chebyshev's Inequality)

Let X be a random variable. Then for any $t > 0$, we have

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \ \geq \ t] \leq \frac{\mathbf{Var}[X]}{t^2}$$

Proof. We first note that

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \geq t] = \mathbf{Pr}[(X - \mathbf{Exp}[X])^2 \geq t^2]$$

Then we notice that $D := (X - \mathbf{Exp}[X])^2$ is a non-negative random variable, and therefore we can apply Markov's inequality on it to get

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \geq t] = \mathbf{Pr}[D \geq t^2] \leq \frac{\mathbf{Exp}[D]}{t^2} = \frac{\mathbf{Var}[X]}{t^2}$$

□

Theorem 3. A useful corollary to the above, and one which is often used as rule of thumb, is obtained by setting $t = c\sigma(X)$ for some $c \geq 0$. One gets,

$$\mathbf{Pr}[|X - \mathbf{Exp}[X]| \ \geq \ c\sigma(X)] \leq \frac{1}{c^2}$$

Proof. When $t = c\sigma(X)$ is substituted in Chebyshev's inequality, one gets the RHS in the above corollary by reminding oneself that $\sigma(X) = \sqrt{\mathbf{Var}(X)}$. □

Example

- Suppose we toss 1000 fair coins. What are the chances that we see more than 600 heads? In this case, let Z be the random variable which evaluates to the number of heads seen in the toss of 1000 coins. We are interested in the question

  $$\mathbf{Pr}[Z \geq 600]?$$

  To evaluate this, we define random variables $X_1, X_2, \ldots, X_{1000}$, where X_i is the indicator random variable for the ith toss; that is, it is defined to be 1 if the ith toss is heads, and it is defined to be 0 if the ith toss is tails. We observe four *crucial* things:

  * $Z = X_1 + X_2 + \cdots + X_{1000}$.
  * $\mathbf{Exp}[X_i] = 0.5$ for all $1 \leq i \leq 1000$. This is because the coins are fair.
  * $X_1, X_2, \ldots, X_{1000}$ are (mutually) *independent*.
  * $\mathbf{Var}[X_i] = 0.25$ (see variance example above – with $p = 0.5$)

Linearity of expectation gives us

$$\mathbf{Exp}[Z] = \sum_{i=1}^{1000} \mathbf{Exp}[X_i] = 1000 \cdot 0.5 = 500$$

The fact that the X_i's are (mutually) independent, allows us to use linearity of variance, to get

$$\mathbf{Var}[Z] = \sum_{i=1}^{1000} \mathbf{Var}[X_i] = 1000 \cdot 0.25 = 250$$

Finally, we can apply Chebyshev's inequality as follows

$$\begin{aligned}
\mathbf{Pr}[Z \geq 600] &= \mathbf{Pr}[Z - 500 \geq 100] && \text{We have subtracted the expectation from both sides} \\
&\leq \mathbf{Pr}[|Z - 500| \geq 100] && \text{if } Z - 500 \geq 100\text{, surely the absolute value is.} \\
&\leq \frac{\mathbf{Var}(Z)}{100^2} && \text{Chebyshev's Inequality} \\
&= \frac{1}{40} && \text{Substituting } \mathbf{Var}[Z] = 250.
\end{aligned}$$

Thus, the chances we see more than 600 heads is *at most* 2.5%.

Remark: *The* true *answer to the question of what is the probability we see more than* 600 *heads is in fact* much, much *lower. The reason is that when a random variable can be written as a sum of mutually independent random variables, then the rule of thumb for the deviations is*

The probability X is more than c standard deviations away is of the order of $e^{-c^2/2}$

The above statement is qualitative rather than quantitative (and therefore I use the term "order of"). But one can see in the above coins example, the standard deviation is $\sqrt{250} \approx 16$. *Thus seeing more than* 100 *heads than the mean is being off by more than* 6 *standard deviations. The chances of this is roughly* $e^{-6^2/2}$ *which is roughly* 1 *in* 100 *million! Way smaller than* 2.5%.

You should use a computer to check it out.

Exercise: *Do the following exercises mimicking the above example.*

- *Suppose every email I get independently is spam with probability* 1%. *I receive* 100 *emails. What is the probability that more than* 7 *of them are spam?*
- *Suppose I roll* 100 *normal dice, and add the sum up. What is the probability that the total sum is less than* 100*?*