Useful Probabilistic Inequalities

DS-563 / CD-543 @ Boston University

Spring 2025 (last updated: 2025-02-04)

The Union Bound

For any probabilistic events $\mathcal{E}_1, \ldots, \mathcal{E}_k$,

$$\Pr(\text{at least one of events } \mathcal{E}_1, \ldots, \mathcal{E}_k \text{ has occured}) \leq \sum_{i=1}^{k} \Pr(\mathcal{E}_i),$$

where $\Pr(\mathcal{E}_i)$ denotes the probability of event $\mathcal{E}_i$.

> In this class, we routinely use the union bound to show that we can avoid a set of bad events with good probability. For instance, consider bad events $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ that can break our algorithm and occur with probability at most $\delta/4$, $\delta/5$, and $\delta/2$, respectively. Then the union bound allows us to say that our algorithm works correctly with probability at least $1 - (\delta/4 + \delta/5 + \delta/2) \geq 1 - \delta$.

Markov's Inequality

Let X be a non-negative random variable with $E[X] < \infty$. For any $a > 0$,

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

> Suppose that a generous stranger leaves an envelope with money in your mailbox every day. If on *average* there is \$100 in the envelope, how often is there at least \$200? Clearly, you cannot find this much in the envelope every day, because then the average would be at least \$200. Can you find this much 51% of the days? Again, the answer is no, because that would imply that the average would be at least $\frac{51}{100} \cdot \$200 = 102$, even if you assume that you get nothing on the remaining 49% of days. Markov's inequality generalizes this type of thinking to give a bound on the probability of a random variable being greater than a specific value.
>
> Exercise: Why is the assumption that the variable is non-negative important in the above reasoning? Would it still hold if the "generous" stranger could take money from you?

Chebyshev's Inequality

Let X be a random variable with finite expectation and variance. For any $a > 0$,

$$\Pr(|X - E[X]| \geq a\sqrt{\mathrm{Var}[X]}) \leq \frac{1}{a^2}.$$

The variance of X, i.e., $\mathrm{Var}[X] = E[(X - E[X])^2]$, is a measure how much on average X diverges from its expectation. If we have a bound on the variance of X, we can bound the probability that X significantly diverges from its expectation. This bound is very useful when X is a sum of other random variables—e.g., $X = \sum_{i=1}^{n} X_i$—that are not fully independent. The standard proof of Chebyshev's inequality is a relatively easy application of Markov's inequality, which uses the fact that $(X - E[X])^2$ is a non-negative variable.

The Chernoff Bound

Let $X_1, \ldots, X_n$ be *independent* random variables taking on values in $[0, 1]$. Let $X = \sum_{i=1}^{n} X_i$ and let $\mu = E[X]$.

For any $\epsilon \in [0, 1]$,

$$\Pr(X \leq (1 - \epsilon)\mu) \leq e^{-\epsilon^2\mu/2},$$

and

$$\Pr(X \geq (1 + \epsilon)\mu) \leq e^{-\epsilon^2\mu/3}.$$

For any $\epsilon \geq 1$,

$$\Pr(X \geq (1 + \epsilon)\mu) \leq e^{-\epsilon\mu/3}.$$

Consider tossing an unbiased coin. Intuitively, you expect that the fraction of both heads and tails will converge to $1/2$ as the number of trials increases. But how fast is it going to happen? This is where the Chernoff bound becomes very useful. As opposed to Chebyshev's inequality, it assumes that the variables are fully independent. This inequality can also be proved via Markov's inequality but the proof is more sophisticated.

Exercise: In the example above, what is the probability that the fraction of heads is at most $2/5$ or at least $3/5$ as a function of n, the number of coin tosses? Set $X_i = 1$ if in the i-th trial the coin comes up heads, and set $X_i = 0$, otherwise.

Bonus: Non-probabilistic Inequalities

For any $x \in \mathbb{R}$, $1 + x \leq e^x$.

Bonus: Notation

Common sets of numbers:

- Natural numbers: $\mathbb{N} = \{0, 1, 2, \ldots\}$
- Integers: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- Real numbers: $\mathbb{R}$

The set of the first n positive integers: For any $n \in \mathbb{N}$, $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$.

Rounding a real number up (aka. the ceiling operator): For any $x \in \mathbb{R}$,

$$\lceil x \rceil \stackrel{\text{def}}{=} \min \{y \in \mathbb{Z} : y \geq x\} .$$

Rounding a real number down (aka. the floor operator): For any $x \in \mathbb{R}$,

$$\lfloor x \rfloor \stackrel{\text{def}}{=} \max \{y \in \mathbb{Z} : y \leq x\} .$$