Useful Probabilistic Inequalities

DS-563 / CD-543 @ Boston University

Fall 2021 (last updated: 11/08/2021)

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.

Chernoff Bound (Multiplicative Concentration)

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 $\delta \in [0, 1]$,

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

and

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

For any $\delta \geq 1$,

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

> 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 when applied to a sum of variables, it assumes independent events. 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.

Collisions (the Birthday Paradox)

We say that there is a *collision* in a set of samples if two of them are identical.

Consider k independent samples x_1, x_2, …, x_k from the uniform distribution on $\{1, \ldots, n\}$. If $k \geq 2\lceil\sqrt{n}\rceil$, then the probability of a collision in this set of samples is at least $1/2$.

> Why? Suppose that there is no collision in the set of the first $\lceil\sqrt{n}\rceil$ samples, i.e., x_1, …, $x_{\lceil\sqrt{n}\rceil}$. Then the probability of any other sample colliding with one of them is at least $\lceil\sqrt{n}\rceil/n \geq 1/\sqrt{n}$. Since the samples are independent, the probability that none of the other $\lceil\sqrt{n}\rceil$ samples collide with them is at most
>
> $$\left(1 - \frac{1}{\sqrt{n}}\right)^{\lceil\sqrt{n}\rceil} \leq e^{-\frac{1}{\sqrt{n}}\cdot\sqrt{n}} = e^{-1} < 1/2.$$
>
> Note 1: It can be showed that the uniform distribution minimizes the probability of a collision, so this bound holds for any distribution, not just the uniform distribution.
>
> Note 2: This problem is referred to as the *birthday paradox*. If one performs the exact computation then a set of 23 people suffices to find a pair with the same birthday with probability more than $1/2$. This may seem counterintuitive, because that's much less than 365, the number of days in a typical year.

Consider k independent samples x_1, x_2, …, x_k from the uniform distribution on $\{1, \ldots, n\}$. For any $p \in [0, 1]$, if $k < \sqrt{2np}$, the probability of seeing a collision is less than p.

> Why? For each pair x_i and x_j, the probability that they are identical, i.e., collide, is $\frac{1}{n}$. Hence the expected number of identical pairs of samples is $\binom{k}{2} \cdot \frac{1}{n} < \frac{k^2}{2n}$. By Markov's inequality, the probability that at least one pair of samples is identical, which is equivalent to having a collision in the set of samples, is at most $\frac{k^2}{2n}$. If $k < \sqrt{2np}$, this is less than p.
>
> In particular, this implies that if we want to see a pair of identical elements drawn with constant probability, we need $\Omega(\sqrt{n})$ samples, i.e., the asymptotic behavior of the previous bound is tight.

Hoeffding's Inequality

Let $X_1, \ldots, X_n$ be *independent* random variables such that each $X_i \in [a_i, b_i]$. For any $t \geq 0$,

$$\Pr\left(\left|\sum_{i=1}^{n} X_i - E[X_i]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

The scenario in which this inequality is most useful in this course is the case of X_i's being indicator variables, or more generally, $X_i \in [0,1]$ for all $i \in [n]$. In this case, the inequality becomes

$$\Pr\left(\left|\sum_{i=1}^{n} (X_i - E[X_i])\right| \geq t\right) \leq 2\exp\left(-2t^2/n\right)$$

for any $t \geq 0$. Alternately, we can write it as

$$\Pr\left(\left|\sum_{i=1}^{n} (X_i - E[X_i])\right| \geq \epsilon n\right) \leq 2\exp\left(-2\epsilon^2 n\right)$$

for any $\epsilon \geq 0$. This should look very familiar to the Chernoff bound, and in fact, in our last homework, we prove a weaker version of this inequality, using the Chernoff bound. The additive bound in Hoeffding's inequality is sometimes more convinient than the multiplicative bound in the Chernoff bound.

Bonus: Non-probabilistic Inequalities

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

More inequalities and other useful info may be added. Stay tuned!