Tips, Tricks, and Techniques for Theoretical Computer Science

Updated: 2015-02-04
Contributors:

- *Thibaut Horel*
- *Bo Waggoner*

Contents

1 Non-Probabilistic Inequalities and Approximations

Exponential function. For all x,

$$1 + x \leq e^x.$$

Easily following are *e.g.* $1 - x \leq e^{-x}$, or $(1+x)^c \leq e^{cx}$, or $\left(1 + \frac{1}{x}\right)^c \leq e^{c/x}$, etc.
See also the Taylor series for e^x.

Logarithm. For all $x > -1$,

$$x - \frac{x^2}{2} \leq \ln(1+x) \leq x.$$

You can push this as far as you want with the Taylor expansion, *e.g.*

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \leq \ln(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}.$$

Bernoulli's Inequality. For all $x \geq -1$, and $n \leq 0$ or $n \geq 1$,

$$1 + xn \leq (1+x)^n.$$

For $0 < n < 1$, the inequality is reversed.
See also the Binomial expansion of $(1+x)^n$ when n is an integer.

Stirling's Approximation for the factorial. The factorial satisfies

$$\left(\frac{n}{e}\right)^n \leq n! \leq n^n.$$

As $n \to \infty$, Stirling's approximation says that

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

This is quite tight; in fact we have[1]

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

Binomial coefficients. The binomial coefficient "n choose k" is

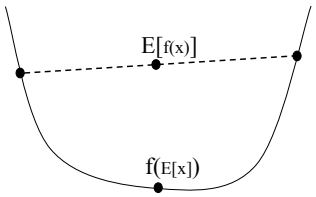
$$\binom{n}{k} = \frac{n!}{(n-k)!k!},$$

and we have

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k.$$

Jensen's Inequality. Suppose f is *convex*: for $\alpha \in (0,1)$, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$. Then for any random variable X,

$$f\left(\mathbb{E}\, X\right) \leq \mathbb{E}\, f(X).$$



In particular, for positive $\{a_i\}$,

$$f\left(\frac{\sum a_i x_i}{\sum a_i}\right) \leq \frac{\sum a_i f(x_i)}{\sum a_i}.$$

For concave functions, all inequalities are reversed.

2 Probabilistic Inequalities and Bounds

Union Bound. For any events $A_1, A_2, \ldots$ (no matter how correlated),

$$\Pr[A_1 \text{ or } A_2 \text{ or } \cdots] \leq \Pr[A_1] + \Pr[A_2] + \cdots .$$

If each A_i has probability p, and there are n of them, then the union bound gives np. If you think they behave approximately independently, then the true probability should be about $1 - (1-p)^n \approx np - O\left((np)^2\right)$. (Using that the Binomial expansion of $(1-p)^n$ is $1 - np + \binom{n}{2}p^2 - \ldots.$)

Markov's Inequality. Let X be a nonnegative real-valued random variable. Then

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

This is especialy useful when both quantities are very small, *e.g.* $\mathbb{E}[X] \to 0$ and we want to bound $\Pr[X \geq 1]$.

Chebyshev's Inequality. Let Y be a real-valued random variable. By applying Markov's to the variable $X = |Y - \mathbb{E}[Y]|^2$, we can get

$$\Pr\left[|Y - E[Y]| \geq b\right] \leq \frac{\mathsf{Var}(Y)}{b^2}.$$

Chernoff Bound for Binomials. Let $X \sim \text{Binomial}(m, p)$ (that is, the number of heads in m independent coin flips with probability p each). Then

$$\Pr[X \leq k] \leq e^{-(mp-k)^2/2mp}.$$

(Of course, mp is the expected number of heads.) Put another way,

$$\Pr[X \leq mp - c\sqrt{mp}] \leq e^{-c^2/2}.$$

You can get a tail bound both above and below: For $k \leq mp$,

$$\Pr[|X - mp| \geq k] \leq 2e^{-k^2/3mp}.$$

A useful reference is Mitzenmacher and Upfal [2].

Hoeffding's Inequality. Essentially a generalization of the above. Let $X_1, \ldots, X_m$ be i.i.d. with each X_i supported on an interval of size b_i; let $S = \sum_i X_i$. Then

$$\Pr\left[|S - \mathbb{E}[S]| \geq k\right] \leq 2e^{-2k^2/\sum_i b_i^2}.$$

Tail bounds in terms of δ. A useful restatement of Hoeffding's is as follows. Let each $b_i = 1$ for simplicity. If we let $k = |S - \mathbb{E}[S]|$, then with probability at least $1 - \delta$,

$$k \leq \sqrt{\frac{m}{2} \ln (2/\delta)}.$$

Such rephrasing can come from any Chernoff-style tail bound and is common in *e.g.* PAC learning.

Chernoff+Union and $\log(n)$. Suppose (for concreteness) we have n Binomials(m, p) and we want to claim that with probability $1 - \delta$, all of them are at most a distance k from their expectation. We can show (notice the new factor of $\log(n)$)

$$k \leq \sqrt{\frac{m}{2} \ln (n/\delta)}$$

because by Chernoff or Hoeffding, each of the n Binomials is within k of its expectation with probability at least $1 - \frac{\delta}{n}$, so by a union bound over the n of them, the probability that any one differs by more than k is bounded by δ.

Note we did not need independence for the union bound. Because of this phenomenon, one often sees the phrasing that a union bound "adds a factor of $\log(n)$".

3 Geometric and Random Phenomena

High-dimensional Cubes. The unit hypercube in $\mathbb{R}^d$ has vertices $\{0,1\}^d$. It has volume 1, but the distance between two opposite vertices (*e.g.* $(0,\ldots,0)$ and $(1,\ldots,1)$) is $\sqrt{d} \to \infty$ as d increases. It is often helpful to visualize the "Boolean hypercube" (the set of vertices of the hypercube) as a sequence or stack of horizontal layers, where each horizontal "slice" is the set of vertices that have k coordinates equal to 1 and $d-k$ coordinates equal to 0, with the "top" ($k=0$) layer containing only $(0,\ldots,0)$ and the "bottom" ($k=d$) layer containing only $(1,\ldots,1)$; the middle layer contains $\binom{d}{2}$ vertices.

High-dimensional Spheres. The unit sphere in $\mathbb{R}^d$ is the set of points at Euclidean distance one from the origin. The volume of the enclosed ball is $\frac{\pi^{d/2}}{\Gamma(1+d/2)}$, where Γ is the generalization of the factorial function to real numbers with $\Gamma(1+x) = x!$ if x is an integer. In particular, the volume approaches zero as $d \to \infty$, although the radius is a constant 1.
A sphere of radius 0.5 centered in the unit cube will touch the center of every face of the cube, yet encloses a volume rapidly approaching zero as d grows (fills almost none of the cube). It may be helpful to visualize the d-dimensional sphere as a "spiky" body with little volume but reaching out in every dimension.

The "Spherical Shell" in High Dimensions. For random vectors with independent coordinates, we often expect concentration in a spherical "shell" at a certain distance from the origin. For instance, suppose we choose a point in $\mathbb{R}^d$ by picking each coordinate X_i in $\{0,1\}$ uniformly and independently. The squared distance to the origin is $\sum_{i=1}^d X_i^2 = \sum_{i=1}^d X_i$, which by the Chernoff bound for Binomials is highly concentrated around $\frac{d}{2}$; in other words, the distance to the origin is concentrated near $\sqrt{d/2}$, which is to say most of the probability lies in a spherical shell.

4 Proof Techniques

Iterated Expectations. *The expected value of X is the expected value, over all values of Y, of the expected value of X given Y.*

$$\mathop{\mathbb{E}}_{X} X = \mathop{\mathbb{E}}_{Y}\left[\mathop{\mathbb{E}}_{X|Y} X\right].$$

This allows computing the expected value of X "indirectly" by marginalizing over Y.

Yao's Principle. *The best deterministic algorithm for a fixed input distribution beats any randomized algorithm on a worst-case input.* Let $\mathcal{A}$ be a randomized algorithm (that is, distribution over deterministic algorithms) and let $\mathcal{X}$ be a distribution over inputs. Then

$$\max_{\text{deterministic algos } a} \mathbb{E}\,\text{performance}(a, \mathcal{X}) \geq \min_{\text{inputs } x} \mathbb{E}\,\text{performance}(\mathcal{A}, x).$$

This is good for showing lower bounds, like "no randomized algorithm has an approximation factor better than c". To prove this, you can construct a distribution over inputs and show that every deterministic algorithm does worse than c on this distribution.

Principle of Deferred Decisions. If you have a randomized algorithm or are *e.g.* building a randomized graph, avoid constructing or reasoning about realizations of a particular piece until your algorithm/analysis touches it. For example, when traversing a random graph, you don't need to reason about the probability of all possible realized graphs, just realizations of the nodes and edges your traversal touches.

References

[1] Herbert Robbins, *A Remark on Stirling's Formula*, The American Mathematical Monthly, 1955.

[2] Michael Mitzenmacher and Eli Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.