Part II

October 18, 2000

6 A Surprising Encounter (of the First Kind?)

(Random) permutations play an essential role in randomized algorithms, so let us take a closer look and recapitulate some simple facts. Recall that a permutation of a set A is simply a bijective mapping $A \to A$. For example, the mapping

$$a \mapsto b, \quad b \mapsto d, \quad c \mapsto a, \quad d \mapsto c$$

is a permutation of the set $\{a, b, c, d\}$. A convenient way of displaying such a permutation π of a finite set $A = \{a_1, a_2, \ldots, a_n\}$ is the *standard notation* (or two row form)

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ \pi(a_1) & \pi(a_2) & \ldots & \pi(a_n) \end{pmatrix} .$$

If there is some natural ordering on the elements in A, as it is the case for $A = \{1..n\}$, then we skip the first row (implicitly assumed to be present in that natural ordering), and we obtain the *linear notation*

$$(\pi(1), \pi(2), \ldots, \pi(n)) \in \{1..n\}^{\underline{n}} .$$

So, we often think of (and use) permutations as orderings, but we should not forget that they are mappings, that can be composed, have inverses, ..., form a group, etc.

A *cycle* in a permutation $\pi \in \mathcal{S}_n$ is a sequence $(i_0, i_1, \ldots, i_{k-1})$ of distinct elements with $\pi(i_j) = i_{j+1 \bmod k}$. Every element $i \in \{1..n\}$ appears in a cycle, that is, the sequence $(i, \pi(i), \pi(\pi(i)), \ldots)$ that terminates just before the second appearance of i. We consider two cycles identical, if they can be obtained form each other by a cyclic shift. Consequently, every element appears in a unique cycle, and the cycles partition $\{1..n\}$. For example, the permutation $(4, 2, 5, 1, 3)$ has cycles $(1, 4)$, (2), and $(3, 5)$. A sequence of all cycles of a permutation is called representation in *cycle notation*, where we put no restriction on the order of cycles, nor where the individual cycles begin. In our example, we could write $(1, 4)(2)(3, 5)$ or $(2)(4, 1)(3, 5)$, among others. Here are all permutations in $\mathcal{S}_3$, their cycles, and the number of cycles.

permutation	cycle notation	number of cycles
1 2 3	(1)(2)(3)	3
1 3 2	(1)(2 3)	2
2 1 3	(1 2)(3)	2
2 3 1	(1 2 3)	1
3 1 2	(1 3 2)	1
3 2 1	(1 3)(2)	2

What is the expected number of cycles of a random permutation in $\mathcal{S}_n$? The obvious route would suggest to count the number of permutations in $\mathcal{S}_n$ with k cycles. In $\mathcal{S}_3$, we get 2 permutations with 1 cycle, 3 with 2, and 1 with 3 – gives an expected number of $1 \times \frac{2}{6} + 2 \times \frac{3}{6} + 3 \times \frac{1}{6} = \frac{11}{6}$ cycles. '$\frac{11}{6}$' sounds familiar, wasn't that H_3? And as we think about it, the table above for $\mathcal{S}_3$ has some similarities with the corresponding table, where we counted the number of left-to-right minima. Indeed, there are 2 permutations with 1 cycle, just like there were 2 permutations with 1 left-to-right minimum, etc. But, a second more careful inspection reveals that the number of cycles of a permutations has little to do with its number of left-to-right minima. $(3, 2, 1)$ has three left-to-right minima, and but it has two cycles.

If you insist, you may still check $\mathcal{S}_4$, and see that it is again the same pattern. So is there a more subtle connection?

Here is how you can show that the number of permutations with k cycles equals the number of permutations with k left-to-right minima. Let us go back to the cycle notation. It leaves some freedom, which we do not only enjoy; it has the disadvantage, that it is not immediately clear whether two permutations in cycle notation are the same. So it is usually agreed upon that every cycle starts with its smallest element, and the cycles are sorted according to their first elements – in ascending order. So $(4, 2, 5, 1, 3)$ will be represented as $(1, 4)(2)(3, 5)$.

Why not descending as: $(3, 5)(2)(1, 4)$? That is, we start with the cycle with the largest smallest element, and so on. We are purists (at least for a moment) and discover that the parenthesis may be omitted. When a new cycle starts, then it starts with an element smaller than all elements in the previous cycle, smaller than all elements so far – a so-called left-to-right minimum! Here we go: $(3, 5, 2, 1, 4)$ represents $(3, 5)(2)(1, 4)$ (and thus $(4, 2, 5, 1, 3)$) in that we agree on starting a new cycle at each left-to-right minimum. This is the desired bijection from $\mathcal{S}_n$ to $\mathcal{S}_n$ (a permutation of $\mathcal{S}_n$) which maps a permutation with k cycles to a permutation with k left-to-right minima.

Hence, once more, nothing really new. The expected number of cycles of a random permutation in $\mathcal{S}_n$ is H_n.

The number of permutations in $\mathcal{S}_n$ with k cycles, $k, n \in \mathbf{N}_0$, is important enough to have a dedicated notation: $\left[{n \atop k}\right]$, the *Stirling cycle number (of the first kind)*; say "n cycle k". It can be recursively defined by

$$\begin{aligned} \left[{0 \atop 0}\right] &= 1, \\ \left[{n \atop 0}\right] &= 0 \quad \text{for } n > 0, \\ \left[{0 \atop k}\right] &= 0 \quad \text{for } k > 0, \text{ and} \\ \left[{n \atop k}\right] &= (n-1)\left[{n-1 \atop k}\right] + \left[{n-1 \atop k-1}\right] \quad \text{for } n, k > 0 . \end{aligned}$$

Note the similarity to the recurrence for binomial coefficients,

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

and how $\sum_{i=0}^{n} \left[{n \atop i}\right] = n!$ relates to $\sum_{i=0}^{n} \binom{n}{i} = 2^n$. You may even (and should) remember, that the latter identity is just a special case of $\sum_{i=0}^{n} \binom{n}{i} x^i = (1+x)^n$. The counterpart for Stirling numbers of the first kind reads

$$\sum_{i=0}^{n} \left[{n \atop i}\right] x^i = \underbrace{x(x+1)\cdots(x+n-1)}_{n \text{ factors}} , \quad \text{for } n \in \mathbf{N}_0 \text{ and } x \in \mathbf{R}. \tag{9}$$

More about cycles and other features of permutations and their expectations in [4].

Exercise 6.1 *Prove the recurrence*

$$\left[{n \atop k}\right] = (n-1)\left[{n-1 \atop k}\right] + \left[{n-1 \atop k-1}\right] .$$

Do it once reading $\left[{n \atop k}\right]$ *as the number of permutations in* $\mathcal{S}_n$ *with* k *cycles, and once reading it as the number of permutations in* $\mathcal{S}_n$ *with* k *left-to-right minima.*

Exercise 6.2 *Determine*

$$\left[{n \atop 1}\right], \left[{n \atop 2}\right], \left[{n \atop n-1}\right], \text{ and } \left[{n \atop n}\right], n \in \mathbf{N}.$$

Exercise 6.3 Sums of Stirling Cycle Numbers
What can you say about

$$\sum_{i=0}^{n} i \left[{n \atop i}\right], \quad \sum_{i=0}^{n} i^2 \left[{n \atop i}\right], \text{ and } \sum_{i=0}^{n} 2^i \left[{n \atop i}\right] ?$$

Exercise 6.4 *Recall the game we analyzed in the section on left-to-right minima. Relate the probability, that such a game starting with* n *takes* k *rounds on one hand, to the number of permutations with a given number of cycles on the other hand.*

Exercise 6.5 $n \in \mathbf{N}$. *For a random permutation* $\pi \in \mathcal{S}_n$ *consider the cycle for which the smallest number is the largest among all cycles in* π. *What is its expected length? What can you say about the expected length of the longest cycle? What about the expected length of the shortest cycle?*

Exercise 6.6 $n \in \mathbf{N}$. *Given* $i \in \{1..n\}$, *what is the expected length of the cycle containing* i *in a random permutation uniformly chosen from* $\mathcal{S}_n$? *Given* $\{i, j\} \in \binom{\{1..n\}}{2}$, *what is the probability that* i *and* j *appear in the same cycle of a random permutation in* $\mathcal{S}_n$?

Exercise 6.7 *We know that for* $n \in \mathbf{N}$, $\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0$. *This can be interpreted as the fact that a nonempty set has the same number of subsets of even cardinality as there are subsets of odd cardinality. Is the corresponding fact true for permutations with an even or odd number of cycles? That is, is it true that*

$$\sum_{i=0}^{n} (-1)^i \left[{n \atop i}\right] = 0 \text{ ?}$$

Exercise 6.8 Estimating the Stirling Cycle Numbers
For $i, n \in \mathbf{N}_0$, $i \le n$, *show that*

$$\left[{n+1 \atop i+1}\right] \le \frac{n!}{i!} (H_n)^i .$$

Exercise 6.9 Increasing Subsequences[6]
Given a permutation $(a_1, a_2, \ldots, a_n)$ *in* $\mathcal{S}_n$, *we call a subsequence*

$$(a_{i_1}, a_{i_2}, \ldots, a_{i_k}) ,$$

$k \in \mathbf{N}_0$, $1 \le i_1 < i_2 < \cdots < i_k \le n$, *an* increasing subsequence, *if*

$$a_{i_1} < a_{i_2} < \cdots < a_{i_k} .$$

For example, $(5\ 1\ 2\ 4\ 3)$ *has the following increasing subsequences*

$$(), (5), (1), (2), (4), (3), (1\ 2), (1\ 4), (1\ 3), (2\ 4), (2\ 3), (1\ 2\ 4), (1\ 2\ 3) .$$

What can you say about the expected number of increasing subsequences of a random permutation in $\mathcal{S}_n$? *What can you say about the length of the longest increasing subsequence?*

7 Coupon Collector

$n \in \mathbf{N}$. Given a set A of n distinct real numbers, the following procedure[7] makes a humble attempt to find the minimum number in A. The statement '$a \leftarrow_{\text{random}} A$;' assigns to a a random element uniformly chosen from A.

$m \leftarrow \infty$;
forever do
  $a \leftarrow_{\text{random}} A$;
  if $a < m$ **then**
    $m \leftarrow a$;

The procedure never knows when it succeeded in finding the minimum. But what is the expected number of steps until m holds $\min A$? What is the probability that m does not hold $\min A$ after k iterations? How long will it take until a was assigned every value in A at least once? What is the expected number of distinct values that a was assigned to within the first k iterations? That's enough to think about for one section (exercises excluded).

We simplify notation by assuming $A = \{1..n\}$. Then the values assigned to a in the first k rounds is a sequence in $\{1..n\}^k$.

[6] A bit out of context, but anyways.

[7] You may find this and previous procedures quite ridiculous in the sense that they pretend to be stupid. However, they allow us to investigate basic methods for analyzing random processes, and, in fact, we will see that identical or very similar structures appear in 'real' programs (– whatever that is).

Waiting for the minimum. m is assigned the minimum when 1 is assigned to a for the first time. In each iteration there is a probability of $\frac{1}{n}$ for that event, and $1 - \frac{1}{n}$ for the complementary event. The waiting time is perfectly modeled by the coin flipping example with probability $p = \frac{1}{n}$ for head to appear. Let W be the random variable, such that round W exhibits the first encounter of 1. Then

$$\Pr(W = i) = (1-p)^{i-1}p, \quad i \in \mathbf{N}$$

and

$$E(W) = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}$$

W is said to have the *geometric distribution with parameter* p. An alternative way of deriving its expectation is as follows. Let I be the indicator variable for the first number chosen to be 1. Then

$$\begin{aligned} E(W) &= \overbrace{E(W \mid I=1)}^{1}\overbrace{\Pr(I=1)}^{p} + \overbrace{E(W \mid I=0)}^{1+E(W)}\overbrace{\Pr(I=0)}^{1-p} \\ &= (1-p)\,E(W) + 1 \end{aligned}$$

which yields $E(W) = \frac{1}{p}$ as before. Here we used that the experiment underlying W is 'memoryless'[8] and so $E(W \mid W \neq 1) = 1 + E(W)$.

Summing up, we expect to wait $\frac{1}{p} = n$ iterations until m holds the minimum of A. The probability that we have not seen the minimum in the first k iterations is $(1-p)^k$. In particular, for $k = n$, this gives $\left(1 - \frac{1}{n}\right)^n < \frac{1}{e}$.

Distinct values in k iterations. Fix $k \in \mathbf{N}$. Let X be the random variable for the number of distinct values that have been assigned to a in the first k rounds. The right set-up and linearity of expectation make this an easy problem. Let X_i, $i \in \{1..n\}$, be an indicator variable for the event that i has *never* been assigned in the first k rounds. Then

$$X = n - X_1 - X_2 - \cdots - X_n ,$$

$\Pr(X_1 = 1) = (1-p)^k$ we had just derived, and, clearly, that's the same for all X_i. It follows that

$$E(X) = n - n\left(1 - \frac{1}{n}\right)^k .$$

For $k = n$, we expect to see roughly $n(1 - \frac{1}{e}) = 0.63_{...}n$ distinct elements. But even for $k = \lfloor n \ln(n/c) \rfloor$, we expect to see only roughly $n - c$ distinct elements. Here we made a rough calculation

$$\left(1 - \frac{1}{n}\right)^{\lfloor n \ln(n/c) \rfloor} \approx e^{-\ln(n/c)} = \frac{c}{n} .$$

[8] The fact happily ignored by the gambler in a casino, who bets on red, since there was a series of ten blacks.

Coupon collector. What is the expectation for the number Y of rounds until all numbers have appeared at least once[9]? As in many situations before and to come, the right twist will make this quite simple to analyze (and at this point you might want to think some time about it before you see the solution).

Think ...

Of course, you knew what to do[10]. Split the process into n phases, where phase i starts after we have seen $i-1$ distinct numbers, and stops when a new number different from all previous ones is encountered. The number of rounds in phase i is denoted by the random variable Y_i, $i \in \{1..n\}$.

$$Y = Y_1 + Y_2 + \cdots + Y_n .$$

If you have absorbed its definition then you will agree that $Y_1 = 1$, always. While waiting for the ith new number, we have a success probability of $p_i := \frac{n-i+1}{n}$ in each round – until it happens. That is, Y_i has geometric distribution with parameter p_i. Thus, $E(Y_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$. We sum up these expectations for $i = 1, 2, \ldots, n$ to obtain

$$E(Y) = \sum_{i=1}^{n} \frac{n}{n-i+1} = nH_n .$$

Note that the Y_i are mutually independent; the X_i in the previous analysis have not been, though.

It is perhaps surprising, that the min-finding procedure is expected to succeed in n rounds much before we expect to have seen all data.

Exercise 7.1 *Prove that a random variable X with the geometric distribution satisfies*

$$\Pr(X = k + i \mid X > i) = \Pr(X = k) \quad \textit{for } i, k \in \mathbf{N},$$

which expresses the 'memorylessness' of the geometric distribution.

Exercise 7.2 *For the random variable Y for the waiting time of the coupon collector, determine $E(X^2)$ and apply Chebyshev's Inequality.*

Exercise 7.3 $n \in \mathbf{N}$. *Let W_i, $i \in \{1..n\}$, be the number of iterations we have to wait until the number i is first assigned to a in the procedure at the beginning of this section. What is $E(W_i)$, $i \in \{1..n\}$ and $E\left(\max_{i \in \{1..n\}} W_i\right)$?*

Exercise 7.4 *Draw a square (about $10\text{cm} \times 10\text{cm}$) on a piece of paper and place 25 random points in it. Do it now!*

....

Now subdivide the square into 25 equal size subsquares, and count the number of empty subsquares. What do you get, and what would be the expected number, if the points were indeed random?

[9] The imaginative reader is invited to see here a coupon collector collecting coupons or such like.

[10] Lucky you, I didn't until I saw the solution!

Exercise 7.5 *What is the expected number of distinct values that m gets assigned to in our little procedure of the section?*

Exercise 7.6 *Your exercises. Remember, I asked you to invent your own exercises!*

8 Chernoff Bounds

Sometimes a random variable X can be written as a sum

$$X = X_1 + X_2 + \cdots + X_n$$

of mutually independent variables X_i. Examples we have met were the number of left-to-right minima, the waiting time of the coupon collector, or the signed difference between the number of heads and tails in a sequence of n coin flips. Here is a lemma for the latter example with a fair coin.

Lemma 8.1 *Let $X = \sum_{i=1}^{n} X_i$ where the X_i's are mutually independent $\{-1,+1\}$-valued random variables with $\Pr(X_i = +1) = \Pr(X_i = -1) = \frac{1}{2}$. Then*

$$\Pr(X \geq \lambda) < \mathsf{e}^{-\lambda^2/(2n)} \quad \textit{for any } \lambda \in \mathbf{R}^{+}.$$

Proof[11] For $t \in \mathbf{R}^{+}$ and $i \in \{1..n\}$, $\mathrm{E}\left(\mathsf{e}^{tX_i}\right) = \frac{1}{2}\left(\mathsf{e}^{t} + \mathsf{e}^{-t}\right) < \mathsf{e}^{t^2/2}$. In order to justify the inequality, we investigate the Taylor series of the terms[12] involved.

$$\begin{aligned}
\frac{1}{2}\left(\mathsf{e}^{t} + \mathsf{e}^{-t}\right) &= \frac{1}{2}\sum_{i=0}^{\infty}\left(\frac{t^i}{i!} + \frac{(-t)^i}{i!}\right) \\
&= \frac{1}{2}\sum_{i=0}^{\infty} 2\frac{t^{2i}}{(2i)!} \\
&< \sum_{i=0}^{\infty}\frac{t^{2i}}{2^i\, i!} \qquad \text{since } \tfrac{1}{(2i)!} \leq \tfrac{1}{2^i\, i!}\text{, strict for } i > 1 \\
&= \mathsf{e}^{t^2/2}
\end{aligned}$$

Mutual independence of the X_i's allows the following estimate.

$$\mathrm{E}\left(\mathsf{e}^{tX}\right) = \mathrm{E}\left(\prod_{i=1}^{n}\mathsf{e}^{tX_i}\right) = \prod_{i=1}^{n}\mathrm{E}\left(\mathsf{e}^{tX_i}\right) < \prod_{i=1}^{n}\mathsf{e}^{t^2/2} = \mathsf{e}^{t^2 n/2}.$$

It is time for Markov's Inequality.

$$\Pr(X \geq \lambda) = \Pr\left(\mathsf{e}^{tX} \geq \mathsf{e}^{t\lambda}\right) \leq \frac{\mathrm{E}\left(\mathsf{e}^{tX}\right)}{\mathsf{e}^{t\lambda}} < \mathsf{e}^{t^2 n/2 - t\lambda},$$

[11] Almost verbatim from [6].

[12] $\left(\mathsf{e}^{t} + \mathsf{e}^{-t}\right)/2$, also denoted as $\cosh(t)$, *hyperbolic cosine*.

for all $t \in \mathbf{R}^{+}$. The parameter t can be chosen so that $\frac{t^2 n}{2} - t\lambda$ is as small as possible. This is attained for $t = \frac{\lambda}{n}$, which yields the statement of the lemma. ◻

The lemma is perhaps more transparent in the form

$$\Pr\left(X \geq \delta\sqrt{n}\right) < \mathsf{e}^{-\delta^2/2} \quad \text{for any } \delta \in \mathbf{R}^{+}. \tag{10}$$

For example, when we toss a fair coin n times, then the probability of the number of heads exceeding the number of tails by $\sqrt{n}$ or more is at most $e^{-1/2} = 0.61\ldots$. If we push the threshold to $\sqrt{2n\ln n}$, then the probability of exceeding is at most $\frac{1}{n}$. Because of the symmetry of X about 0, we have

$$\Pr\left(|X| \geq \delta\sqrt{n}\right) < 2\mathsf{e}^{-\delta^2/2} \quad \text{for any } \delta \in \mathbf{R}^{+}. \tag{11}$$

We show a typical employment of the lemma, so that we can appreciate its potentials.

Regular partitions of regular hypergraphs. A *hypergraph* $G = (V, E)$, $E \subseteq 2^V$ is r-*regular*, if all its vertices have degree r, i.e. each vertex is contained in exactly r hyperedges – edges for short from now on. We want to partition the edge set E of an r-regular hypergraph into E_{red} and E_{blue} so that both $G_{\mathrm{red}} = (V, E_{\mathrm{red}})$, the red subgraph, and $G_{\mathrm{blue}} = (V, E_{\mathrm{blue}})$, the blue subgraph, are as regular as possible. Ideally, we want them both to be $\frac{r}{2}$-regular. That's odd to ask for r odd, but even for r even that is not always achievable: Try your luck with the ordinary graph of an odd cycle.

Theorem 8.1 $r, n \in \mathbf{N}$. *Every r-regular hypergraph with n vertices can be partitioned into a red and a blue subgraph so that all degrees appearing in both hypergraphs are in*

$$\left\{\left\lceil \frac{r}{2} - \sqrt{\frac{r\ln(2n)}{2}}\right\rceil .. \left\lfloor \frac{r}{2} + \sqrt{\frac{r\ln(2n)}{2}}\right\rfloor\right\}.$$

Proof Assign random colors to the edges of the hypergraph $G = (V, E)$, both colors with probability $\frac{1}{2}$, and mutually independent for all the edges. In fact, we will use -1 and $+1$ for red and blue, respectively. Denote by X_e the value assigned to edge $e \in E$, and for vertex $v \in V$, let

$$X^{(v)} := \sum_{e \ni v} X_e.$$

$\frac{r - X^{(v)}}{2}$ is the degree in the red (-1) subgraph, and $\frac{r + X^{(v)}}{2}$ the degree in the blue $(+1)$ subgraph. If we can show that there is an assignment with $\left|X^{(v)}\right| < \sqrt{2r\ln(2n)}$ for all $v \in V$, we are done.

According to (11),

$$\Pr\left(\left|X^{(v)}\right| \geq \delta\sqrt{r}\right) < 2\mathsf{e}^{-\delta^2/2} \quad \text{for each } v \in V,$$

since each $X^{(v)}$ is the sum of r of the variables X_e. Consequently,

$$\Pr\left(\bigvee_{v \in V}\left(\left|X^{(v)}\right| \geq \delta\sqrt{r}\right)\right) < 2n\mathsf{e}^{-\delta^2/2}.$$

If this probability is less than 1, there remains a positive probability for the complementary event

$$\bigwedge_{v\in V}\left(\left|X^{(v)}\right| < \delta\sqrt{r}\right) \ ,$$

i.e. such an event exists! We choose $\delta = \sqrt{2\ln(2n)}$ and the assertion of the theorem follows. ◻

The theorem demonstrates existence. As we increase δ in the proof, a random coloring will be good with increasing probability.

Estimating sums of binomial coefficients. Can't we determine $\Pr(X \geq \lambda)$ exactly (X as in Lemma 8.1)? X is related to the binomial distribution with parameters n and $\frac{1}{2}$ and it is easy to evaluate $\Pr(X = i)$ for $i \in \mathbf{N}_0$. There are 2^n sequences of $+1$'s and -1's, each appears as $(X_1, X_2, \ldots, X_n)$ with probability $\frac{1}{2^n}$. There are $\binom{n}{j}$ sequences with j $(+1)$'s (and thus $n-j$ (-1)'s), which let X attain the value $j-(n-j) = 2j-n$. Hence, $\Pr(X = 2j-n) = \frac{1}{2^n}\binom{n}{j}$ and

$$\Pr(X \geq 2j-n) = \frac{1}{2^n}\sum_{i=j}^{n}\binom{n}{i}\ , \quad \text{for any } j \in \mathbf{N}_0. \tag{12}$$

Now we know the answer exactly, but we have no clue of what it means. Our findings are not useless, though. We can use them to estimate the sum of the first k binomial coefficients by invoking Lemma 8.1.

$$\sum_{i=0}^{k}\binom{n}{i} = \sum_{i=n-k}^{n}\binom{n}{i} = 2^n\Pr(X \geq 2(n-k)-n) < 2^n\mathsf{e}^{-\frac{(n-2k)^2}{2n}}\ , \tag{13}$$

for $k \in \mathbf{N}_0$, $k < \frac{n}{2}$. Thus, e.g. the sum up to $k = \frac{n-\sqrt{2n\ln n}}{2}$ makes up for less than $\frac{1}{n}$ of the whole sum of binomial coefficients. In contrast, we have $\binom{n}{\lfloor n/2\rfloor} = \Theta(\frac{2^n}{\sqrt{n}})$.

It is instructive to rephrase (13) as a property of the hypercube[13], or of $\{0,1\}$-strings and the Hamming distance[14].

Chernoff bound technique. There is nothing like 'the Chernoff Bound'. This term refers to a technique for establishing good tail estimates for a random variable X that is the sum of independent variables. In order to do so, one analyzes $\mathrm{E}\left(\mathsf{e}^{tX}\right)$, $t \in \mathbf{R}$ and then applies Markov's Inequality, while setting t to the value that gives the strongest possible result. Since

$$\mathrm{E}\left(\mathsf{e}^{tX}\right) = 1 + \frac{t\,\mathrm{E}(X)}{1!} + \frac{t^2\,\mathrm{E}(X^2)}{2!} + \cdots$$

[13]The graph with vertex set $\{0,1\}^n$ and edges between any two such sequences that differ in exactly one position.

[14]The *Hamming distance* between two sequences (strings) in $\{0,1\}^n$ is the number of positions where they differ.

this expectation – parameterized by t – is called the moment generating function of X. For collections of Chernoff bounds see [1, Appendix A], [5, Section 4.1] or [3].

Here is one more example of a Chernoff bound.

Lemma 8.2 *Let X be the sum of a finite number of mutually independent $\{0,1\}$-valued random variables such that $\mu := \mathrm{E}(X)$ is positive. Then*

$$\Pr(X \leq (1-\delta)\mu) < \left(\frac{\mathsf{e}^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} < \mathsf{e}^{-\mu\delta^2/2} \quad \textit{for any } \delta \in \mathbf{R}^{+} \textit{ with } \delta < 1.$$

Before we proceed with the proof, let us briefly discuss the lemma. First note that it is a bound for $\Pr(X \leq \lambda)$ (rather than $\Pr(X \geq \lambda)$) – and beware, in general,

$$\Pr(X \leq (1-\delta)\mu) \neq \Pr(X \geq (1+\delta)\mu)\ .$$

Second, observe that the bound does not refer to the number of variables involved, nor at their individual distributions.

Let us apply the lemma to the random variable X for the number of left-to-right minima in a random permutation in $\mathcal{S}_n$. This variable satisfies the assumptions of the lemma with $\mu = H_n$. We choose $\delta = \frac{1}{2}$ and conclude that

$$\Pr\left(X \leq \frac{1}{2}H_n\right) < \mathsf{e}^{-H_n/8} \leq (n+1)^{-1/8}\ .$$

The bound looks more impressive for variables with a larger expectation.

Proof of Lemma 8.2 Since the estimate is of the form $\Pr(X \leq \lambda)$ rather than $\Pr(X \geq \lambda)$ we switch from X to $-X$ in order to return to familiar terrain. That is, we will use Markov's Inequality as follows for $t \in \mathbf{R}^{+}$.

$$\Pr(X \leq (1-\delta)\mu) = \Pr(-X \geq (\delta-1)\mu) = \Pr\left(\mathsf{e}^{-tX} \geq \mathsf{e}^{t(\delta-1)\mu}\right) \leq \frac{\mathrm{E}\left(\mathsf{e}^{-tX}\right)}{\mathsf{e}^{t(\delta-1)\mu}} \tag{14}$$

Now let $X = \sum_{i=1}^{n} X_i$, $n \in \mathbf{N}$, where the X_i's are independent $\{0,1\}$-valued random variables, and let $p_i := \Pr(X_i = 1)$, $1 \leq i \leq n$. Therefore, $\mu = \sum_{i=1}^{n} p_i$. Now

$$\mathrm{E}\left(\mathsf{e}^{-tX}\right) = \prod_{i=1}^{n}\mathrm{E}\left(\mathsf{e}^{-tX_i}\right) = \prod_{i=1}^{n}\left(1 + p_i(\mathsf{e}^{-t}-1)\right) < \mathsf{e}^{\mu(\mathsf{e}^{-t}-1)} \tag{15}$$

Next we plug the estimate from (15) into (14).

$$\Pr(X \leq (1-\delta)\mu) < \mathsf{e}^{\mu(\mathsf{e}^{-t}-1+t(1-\delta))} \tag{16}$$

The bound is smallest for $t = -\ln(1-\delta)$ when it gives a bound of

$$\mathsf{e}^{\mu((1-\delta)(1-\ln(1-\delta))-1)}$$

which can be rewritten as the first bound of the lemma. For the second simplified form we use the inequality

$$(1-\delta)\ln(1-\delta) > -\delta + \delta^2/2 \quad \text{for } \delta \in \mathbf{R}, 0 < \delta < 1.$$

$\square$

We conclude with the statement of the counterpart of Lemma 8.2, without delivering the proof. But beware, there are many more Chernoff bounds, and when you need one, you might have to use the idea rather than fetching a nicely packed result from the shelf.

Lemma 8.3 *Let X be the sum of a finite number of mutually independent $\{0,1\}$-valued random variables such that $\mu := \mathrm{E}(X)$ is positive. Then*

$$\Pr(X \geq (1+\delta)\mu) < \left(\frac{\mathsf{e}^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad \textit{for any } \delta \in \mathbf{R}^{+}.$$

Exercise 8.1 A Regular Hypergraph
$i, n \in \mathbf{N}$. Consider the hypergraph $G = (\{1..n\}, \binom{\{1..n\}}{i})$. Show that G is r-regular for some r. Which r? Find a good partition into a red and a blue subgraph, both as regular as possible.

Exercise 8.2 *$n \in \mathbf{N}$. Let Y_i, $i \in \{1..n\}$, be the random variable for phase i in the coupon collector analysis. Determine $\mathrm{E}\left(a^{Y_i}\right)$, for $a \in \mathbf{R}$.*

Exercise 8.3 Chernoff Bound for Coupon Collector
Derive a Chernoff bound for the coupon collector exceeding a certain threshold λ.

Exercise 8.4 *Verify that Theorem 8.1 is still valid, if all vertices have degree* at most *r.*

Exercise 8.5 Orthogonal Vectors
$n \in \mathbf{N}$ and $v \in \{-1,+1\}^n$. How many vectors in $\{-1,+1\}^n$ are orthogonal to v?

Exercise 8.6 Close to Orthogonal Vectors
$n \in \mathbf{N}$ and $v \in \{-1,+1\}^n$. How many vectors in $\{-1,+1\}^n$ are close to orthogonal to v? We define close to orthogonal *by requiring an angle in the interval $\left(\frac{\pi}{2}-\varepsilon, \frac{\pi}{2}+\varepsilon\right)$ for some given parameter $\varepsilon \in \mathbf{R}^{+}$.*

Exercise 8.7 Orthogonal to Many Vectors
$n \in \mathbf{N}$. Show that for any set of n vectors in $\{-1,+1\}^n$, there exists a vector that is close to orthogonal to all of them. How close? (See Exercise 8.6 for what we mean by 'close to orthogonal'.)

Exercise 8.8 *$k, n \in \mathbf{N}$. What can you say about*

$$\sum_{i=0}^{k} \begin{bmatrix} n \\ i \end{bmatrix} \ ?$$

Exercise 8.9 Commercial
Your exercise could be placed here – and you were reading it, and the many others who have made it that far.

References

[1] Noga Alon, Joel H. Spencer, *The Probabilistic Method*, Wiley-Interscience (1992).

[2] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, *Concrete Mathematics; A Foundation for Computer Science*, Addison-Wesley (1989).

[3] Torben Hagerup, Christiane Rüb, A guided tour of Chernoff bounds, *Information Processing Letters* **33** (1990) 305-308.

[4] Rainer Kemp, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner Series in Computer Science, (1984).

[5] Rajeev Motwani, Prabhakar Raghavan, *Randomized Algorithms*, Cambridge University Press (1995).

[6] Joel H. Spencer, *Ten Lectures on the Probabilistic Method*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM (1987).

A Glossary of Facts

This is a pretty much unorganized collection of useful facts, which is perhaps boring, so you might want to skip it – but *only if* you know all of this!

Fact A.1

$$1 + x \leq \mathsf{e}^{x} \quad \textit{for } x \in \mathbf{R}, \textit{ strict for } x \neq 0.$$

It is useful for deriving upper bounds for products. For example,

$$\prod_{i=1}^{\infty}\left(1+\frac{1}{i^2}\right) < \prod_{i=1}^{\infty} \mathsf{e}^{1/i^2} = \mathsf{e}^{\sum_{i=1}^{\infty} 1/i^2} = \mathsf{e}^{\pi^2/6} = 5.18\ldots \ . \tag{17}$$

and for $n \in \mathbf{N}$

$$\prod_{i=1}^{n}\left(1+\frac{1}{2i}\right) < \prod_{i=1}^{n} \mathsf{e}^{1/(2i)} = \mathsf{e}^{\sum_{i=1}^{n} 1/(2i)} = \mathsf{e}^{H_n/2} \leq \underbrace{\sqrt{\mathsf{e}n}}_{1.65\ldots\sqrt{n}} \tag{18}$$

Another appearance of the inequality is in estimates of the form[15]

$$\left(1+\frac{x}{n}\right)^{n} \leq \mathsf{e}^{(x/n)n} = \mathsf{e}^{x}, \quad \text{for } n \in \mathbf{N} \text{ and } x \in \mathbf{R}. \tag{19}$$

[15] You are invited to digest the special case $n = 1$! And $x = 1$.

Fact A.2

$$\frac{1}{1-x} \geq \mathsf{e}^x \quad \textit{for } x \in \mathbf{R},\ x < 1, \textit{ strict for } x \neq 0.$$

It delivers lower bounds for products of numbers close to 1. For example, for $n \in \mathbf{N}$

$$\prod_{i=1}^{n}\left(1+\frac{1}{2i}\right) \;>\; \prod_{i=1}^{n} \mathsf{e}^{1/(2i+1)} = \mathsf{e}^{\sum_{i=1}^{n} 1/(2i+1)} \tag{20}$$

$$= \mathsf{e}^{H_{2n+1}-H_n/2-1} \geq \frac{2n+2}{\mathsf{e}\sqrt{\mathsf{e}n}} > \underbrace{\frac{2}{\mathsf{e}\sqrt{\mathsf{e}}}}_{0.45\ldots} \sqrt{n}\,. \tag{21}$$

Fact A.3

$$x+(1-x)\ln(1-x) \geq \frac{1}{2}x^2 \quad \textit{for } x \in \mathbf{R}_0^+,\ x < 1, \textit{ strict for } x \neq 0.$$

Okay, that's a bit special. You are excused, if you don't know that one if I wake you up at 4am.

Fact A.4

$$\mathsf{e}^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} \quad \textit{for } x \in \mathbf{R}.$$

That we knew! But what happens if $(i!)^2$ appears in the denominator?

$$\sum_{i=0}^{\infty} \frac{x^i}{(i!)^2} = \sum_{i=0}^{\infty}\left(\frac{x^{i/2}}{i!}\right)^2 < \left(\sum_{i=0}^{\infty}\frac{x^{i/2}}{i!}\right)^2 = \mathsf{e}^{2\sqrt{x}} \quad \text{for } x \in \mathbf{R}^+. \tag{22}$$

Fact A.5

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i \quad \textit{for } x \in \mathbf{R},\ -1 < x < 1.$$

Take the first derivative on both sides:

$$\frac{1}{(1-x)^2} = \sum_{i=1}^{\infty} i\, x^{i-1} \quad \text{for } x \in \mathbf{R},\ -1 < x < 1. \tag{23}$$

Fact A.6 (Harmonic vs. Geometric vs. Arithmetic Mean) $n \in \mathbf{N}$. *For any* $(a_1, a_2, \ldots, a_n) \in {\mathbf{R}^+}^n$

$$\frac{n}{\sum_{i=1}^{n}\frac{1}{a_i}} \;\leq\; \sqrt[n]{\prod_{i=1}^{n} a_i} \;\leq\; \frac{\sum_{i=1}^{n} a_i}{n}\,;$$

both inequalities strict, unless all a_i the same.

A typical usage is the immediate implication

$$\sum_{i=1}^{n}\frac{1}{a_i} \geq \frac{n^2}{\sum_{i=1}^{n} a_i} \tag{24}$$

Fact A.7 (Cauchy-Schwartz Inequality) $n \in \mathbf{N}$. *For any* $(a_1, a_2, \ldots, a_n) \in \mathbf{R}^n$, $(b_1, b_2, \ldots, b_n) \in \mathbf{R}^n$,

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geq \left(\sum_{i=1}^{n} a_i b_i\right)^2;$$

strict, unless there exists a $\lambda \in \mathbf{R}$ with $a_i = \lambda b_i$ for all $i \in \{1..n\}$.

By setting all $b_i = 1$ we obtain as a special case

$$\frac{\sum_{i=1}^{n} a_i}{n} \leq \sqrt{\frac{\sum_{i=1}^{n} a_i^2}{n}} \quad \text{for any } (a_1, a_2, \ldots, a_n) \in \mathbf{R}^n \tag{25}$$

with equality iff all a_i are the same (arithmetic vs. quadratic mean).

Fact A.8 (Stirling's Formula)

$$n! = \sqrt{2\pi n}\left(\frac{n}{\mathsf{e}}\right)^n\left(1+O(n^{-1})\right),\ \ n \in \mathbf{N}\,.$$

As a variation on the theme

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} = \frac{2\sqrt{\pi n}\left(\frac{2n}{\mathsf{e}}\right)^{2n}\left(1+O(n^{-1})\right)}{2\pi n\left(\frac{n}{\mathsf{e}}\right)^{2n}\left(1+O(n^{-1})\right)} = \frac{1}{\sqrt{\pi n}}2^{2n}\left(1+O(n^{-1})\right)\,. \tag{26}$$

This gives us access to a better estimate for the product from (18) and (20).

$$\prod_{i=1}^{n}\left(1+\frac{1}{2i}\right) \;=\; \prod_{i=1}^{n}\frac{2i+1}{2i} = \frac{(2n+1)!}{(2^n\, n!)^2} \tag{27}$$

$$= \binom{2n}{n}\frac{2n+1}{2^{2n}} = \frac{2n+1}{\sqrt{\pi n}}\left(1+O(n^{-1})\right) \tag{28}$$

$$= \underbrace{\frac{2}{\sqrt{\pi}}}_{1.13\ldots}\sqrt{n}\left(1+O(n^{-1})\right) \tag{29}$$

Fact A.9 (Symmetry and Factorial Expansion of Binomial Coefficients) *For* $i, n \in \mathbf{N}_0$, $i \leq n$,

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

Fact A.10 *For $i, n \in \mathbf{N}$, $0 < i \le n$,*

$$\left(\frac{n}{i}\right)^i \le \binom{n}{i} \le \frac{n^i}{i!} < \left(\frac{\mathsf{e}n}{i}\right)^i .$$

The rightmost inequality is a consequence of

$$n! \ge \left(\frac{n}{\mathsf{e}}\right)^n \qquad \text{for } n \in \mathbf{N}_0\text{, strict for } n \ne 0. \tag{30}$$

Fact A.11 (Binomial Theorem) *For any $n \in \mathbf{N}_0$ and $x, y \in \mathbf{R}$,*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} .$$

For $x = y = 1$ we get

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n \tag{31}$$

as special case. Or, for $x = -1$ and $y = 1$,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0 . \tag{32}$$

Whenever you see an expression of the form $x^n + n\,yx^{n-1}$, remember the Binomial Theorem and estimate

$$x^n + n\,yx^{n-1} \le (x+y)^n \quad \text{for } x, y \in \mathbf{R}^+, n \in \mathbf{N}\text{, strict for } n \ne 1. \tag{33}$$

B More Exercises

Exercise B.1 Generating a Group

Let G be a finite group of order n with neutral element e. For $A \subseteq G$, $\langle A \rangle$ denotes the subgroup of G generated by A. Consider the following procedure.

$A \leftarrow \{e\}$;
forever do
  $a \leftarrow_{\text{random}} G$;
  if $a \notin \langle A \rangle$ **then**
    $A \leftarrow A \cup \{a\}$;

Find estimates for the expected time it takes until A generates G. (Take a look at specific groups as well, e.g. $(\mathbf{Z}_2)^d$, $(\mathbf{Z}_3)^d$, $d \in \mathbf{N}$, or the cyclic groups $\mathbf{Z}_n$, $n \in \mathbf{N}$.)

Exercise B.2 Independence and Conditional Probabilities

Prove or disprove each of the following three statements.

(1) Random variables X and Y on probability space $(\Omega, \Pr)$ are independent iff

$$\Pr(X = x \,|\, Y = y) \;=\; \Pr(X = x)$$

for all $x \in \mathbf{R}$ and all $y \in Y(\Omega^+)$.

(2) $n \in \mathbf{N}$. Random $\{0,1\}$-valued random variables $X_1, X_2, \ldots, X_n$, none of them constant, are mutually independent iff

$$\Pr\left(X_j = 1 \,|\, \bigwedge_{i \in J\setminus\{j\}} (X_i = 1)\right) \;=\; \Pr(X_j = 1)$$

for all $J \subseteq \{1..n\}$ and all $j \in J$.

(3) $n \in \mathbf{N}$. Random $\{0,1\}$-valued random variables $X_1, X_2, \ldots, X_n$, none of them constant, are mutually independent iff for all nonempty $J \subseteq \{1..n\}$ there exists $j \in J$ such that

$$\Pr\left(X_j = 1 \,|\, \bigwedge_{i \in J\setminus\{j\}} (X_i = 1)\right) \;=\; \Pr(X_j = 1) .$$

Exercise B.3 Tail Estimate for Paranoia

Derive tail estimates for the number of comparisons in the procedure of Exercise 3.6 (try also a Chernoff bound).

Exercise B.4 Slow Minimum

Here is another one of those 'stupid' algorithms that find the minimum of a finite nonempty *set A of n reals.*

function FindMin(A)
$a \leftarrow_{\text{random}} A$;
if $\#A = 1$ **then**
    return a;
else
    $x \leftarrow$ FindMin($A \setminus \{a\}$);
    if $a < x$ **then**
        $x \leftarrow$ FindMin(A);
    return x;

Make sure that you understand what the procedure does, and that, indeed, it computes the minimum of A. Note that the procedure invokes a recursive call with the same parameters – a feature that is not acceptable for a deterministic procedure (at least, if termination is desired).

Analyze the expected running time of the procedure. For that purpose count the number of comparisons '$a < x$' performed.

Exercise B.5 Recursion

Consider the function $f(n,k)$ for integers k and n, $0 \leq k \leq n$, recursively defined by $f(k,k) = 0$ for $k \in \mathbf{N}_0$ and

$$f(n,k) = f(n-1,k) + 1 + \frac{1}{n-k} f(n,k+1)\,, \quad \textit{for } 0 \leq k < n.$$

Find a closed form for $f(n,k)$.