Basic Examples of Probabilistic Analysis, Part I

Reading Assignment for PreDoc Course on Randomized Algorithms

Emo Welzl, ETH Zürich

October 18, 2000

This is a collection of basics and basic examples of probabilistic analysis of discrete random variables, very much biased towards what we will need in the course on Randomized Algorithms. Even if you know some of the problems, it is important to understand the methods used for the analysis.

There are many exercises, some easy and some difficult, and what of the two criteria applies may very much depend on the solver. Moreover, beware of the fact that some of the exercises are ill-posed (let it be intentional or because I made a mistake) – that is, the problem statement has to be corrected or made more precise. It is part of the exercise to do so.

Perhaps you won't do all the exercises, but those you do, do them carefully and prepare a clean written solution. Writing is part of a scientist's life (and of a Ph.D. student's, in particular), and I recommend to train this skill on a small scale.

Finally, my experience is that besides doing exercises the best training for understanding some material is to design your own exercises. More than once, by doing so I discovered new results, some of which even led to publications. So go ahead and surprise (and tease) your colleagues and me with exercises (of which you know the solutions) or open problems! This is very much in the tradition of great mathematicians. For example, when discovering a new result, Fermat would not publish the proof, but challenge his colleagues instead. We know that the challenge lasted for quite some time in some instances (and think about it – that's what he is famous for).

1 Notation

Briefly browse through this list. I hope this notational prologue doesn't scare you off; it will probably not contain any big surprises. If so, make sure you learn something about the respective entities!

Doing your exercises, you are not expected to stick to these notations (after all, the parallel course may deviate in details), but make sure that you for yourself agree on some reasonable consistent notation. Actually, on the blackboard, I will use $\mathbb{N}$ for $\mathbf{N}$, $\mathbb{Z}$ for $\mathbf{Z}$, $\mathbb{R}$ for $\mathbf{R}$ etc.

$$\begin{array}{rcll}
\mathbf{N} & := & \{1,2,\ldots\} & \ldots\text{natural numbers} \\
\mathbf{N}_0 & := & \{0,1,2,\ldots\} & \ldots\text{nonnegative integers} \\
\mathbf{Z} & := & \{\ldots,-2,-1,0,1,2,\ldots\} & \ldots\text{integers} \\
\{i..j\} & := & \{n\in\mathbf{Z} \mid i\le n\le j\},\ \ i,j\in\mathbf{Z} & \\
\mathbf{R} & & \ldots\text{real numbers} & \\
\mathbf{R}^+ & := & \{x\in\mathbf{R} \mid x>0\}\} & \ldots\text{positive real numbers} \\
\mathbf{R}^+_0 & := & \{x\in\mathbf{R} \mid x\ge 0\}\} & \ldots\text{nonnegative real numbers}
\end{array}$$

For $n,k\in\mathbf{N}_0$,

$$\begin{array}{rcll}
H_n & := & \displaystyle\sum_{i=1}^{n}\frac{1}{i} = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} & \ldots n\text{th Harmonic number} \\
n! & := & \displaystyle\prod_{i=1}^{n} i = \begin{cases} 1 & \text{if } n=0, \\ n\cdot(n-1)! & \text{if } n>0. \end{cases} & \ldots n \text{ factorial} \\
\displaystyle\binom{n}{k} & := & \dfrac{\prod_{i=0}^{k-1}(n-i)}{k!} & \ldots\text{binomial coefficient} \\
n^{\underline{k}} & := & \displaystyle\prod_{i=0}^{k-1}(n-i) & \ldots\text{falling powers}
\end{array}$$

For a set A and $k\in\mathbf{Z}$,

$$\begin{array}{rl}
\#A & \ldots\text{cardinality of set } A \\
2^A & \ldots\text{set of subsets of } A \text{ (power set)} \\
A^k & \ldots\text{set of sequences over } A \text{ of length } k \\
A^{\underline{k}} & \ldots\text{set of sequences of length } k \text{ of distinct elements over } A \\
\displaystyle\binom{A}{k} & \ldots\text{set of subsets of } A \text{ of cardinality } k
\end{array}$$

Note that for a finite set A, $n=\#A$, and $k\in\mathbf{N}_0$,

$$\#2^A = 2^n,\ \ \#(A^k)=n^k,\ \ \#(A^{\underline{k}})=n^{\underline{k}},\quad\text{and}\quad \#\binom{A}{k}=\binom{n}{k}\,.$$

Here is an important real number, and our notation for logarithms.

$$\mathsf{e} \;:=\; \sum_{i=0}^{\infty}\frac{1}{i!} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = 2.718\ldots$$

or the real number such that $\int_1^{\mathsf{e}}\frac{dt}{t}=1$

$$\begin{array}{rcll}
\log_b & & \ldots\text{logarithm base } b,\ b\in\mathbf{R}^+ & \\
\ln & := & \log_{\mathsf{e}} & \ldots\text{natural logarithm} \\
\lg & := & \log_2 & \ldots\text{binary logarithm}
\end{array}$$

For a statement S that can be true or false,

$$[S] \;:=\; \begin{cases} 1 & \text{if } S \text{ is true,} \\ 0 & \text{if } S \text{ is false.} \end{cases} \qquad \text{...indicator function for statement } S$$

And two extra entries:

$$\min, \max \qquad \text{...minimum and maximum of a set of real numbers,}$$
$$\min \emptyset := \infty, \max \emptyset := -\infty$$
$$\mathcal{S}_n \qquad \text{...set (group) of permutations of } \{1..n\},\ n \in \mathbf{N}$$

2 Discrete Probability

I assume that you have some preknowledge in probability theory, but let us recapitulate the small subset of it which we will need for our purposes. We will employ very concrete probability theory (as opposed to abstract probability theory); we simply use it as a tool. When writing this I borrowed much from chapter 8 in [2], a book to be recommended not only for that chapter.

We restrict ourselves to discrete probability spaces, and we will omit 'discrete' from now on. Roughly speaking, such a probability space consists of a (possibly infinite) set of things that can happen, each of which gets assigned a probability that it happens. This mapping is called probability distribution, since it distributes the value 1 among the things that can happen.

Definition 2.1 (Probability Space) *A* probability space *is a pair* (Ω, Pr) *where* Ω *is a set and* Pr *is a mapping* $\Omega \to \mathbf{R}_0^+$ *such that*

$$\sum_{\omega \in \Omega} \mathrm{Pr}(\omega) = 1 \;.$$

Every subset A *of* Ω *is called an* event, *and the mapping* Pr *is extended to events by setting*

$$\mathrm{Pr}(A) := \sum_{\omega \in A} \mathrm{Pr}(\omega) \;.$$

The elements in Ω *are the* elementary events. *If* Ω *is finite and* $\mathrm{Pr}(\omega) = \frac{1}{\#\Omega}$ *for all* $\omega \in \Omega$*, then* Pr *is called* uniform distribution on Ω. *We use* Ω^+ *for the set of elementary events with positive probability,*

$$\Omega^+ := \{\omega \in \Omega \mid \mathrm{Pr}(\omega) > 0\} \;.$$

Rolling dice. Consider the six sides of a die denoted by

$$D = \left\{\boxed{\texttt{1}}, \boxed{\texttt{2}}, \boxed{\texttt{3}}, \boxed{\texttt{4}}, \boxed{\texttt{5}}, \boxed{\texttt{6}}\right\} \;.$$

D models the top side of the die as it lands on the table in an experiment. We consider fair dice, i.e. $\mathrm{Pr}(d) = \frac{1}{6}$ for all $d \in D$. The pair (D, Pr) is a probability space with uniform distribution.

For example, $D_{\text{even}} = \{\boxed{\texttt{2}}, \boxed{\texttt{4}}, \boxed{\texttt{6}}\}$ is the event of having an even number of spots on the top side; $\mathrm{Pr}(D_{\text{even}}) = 3 \times \frac{1}{6} = \frac{1}{2}$.

Rolling a pair of fair dice is modeled by the set

$$\mathbb{D} := D^2 = \left\{\boxed{\texttt{1}}\boxed{\texttt{1}}, \boxed{\texttt{1}}\boxed{\texttt{2}}, \boxed{\texttt{1}}\boxed{\texttt{3}}, \ldots, \boxed{\texttt{6}}\boxed{\texttt{5}}, \boxed{\texttt{6}}\boxed{\texttt{6}}\right\}$$

of 36 elementary events with the uniform distribution. Note that the two dice are assumed to be distinguishable, say one as the first die, and the other as the second. For the event

$$\mathbb{D}_{=} = \left\{\boxed{\texttt{1}}\boxed{\texttt{1}}, \boxed{\texttt{2}}\boxed{\texttt{2}}, \boxed{\texttt{3}}\boxed{\texttt{3}}, \boxed{\texttt{4}}\boxed{\texttt{4}}, \boxed{\texttt{5}}\boxed{\texttt{5}}, \boxed{\texttt{6}}\boxed{\texttt{6}}\right\}$$

we have $\mathrm{Pr}(\mathbb{D}_{=}) = 6 \times \frac{1}{36} = \frac{1}{6}$, and we say that the probability of having the same number of spots on both dice is $\frac{1}{6}$. The event, $\mathbb{D}_{\neq}$, of having a distinct number of spots on the dice is the event complementary to $\mathrm{Pr}(\mathbb{D}_{=})$; hence, $\mathrm{Pr}\left(\mathbb{D}_{\neq}\right) = 1 - \mathrm{Pr}(\mathbb{D}_{=}) = \frac{5}{6}$. The event $\mathbb{D}_{\neq}$ partitions into the event $\mathbb{D}_{<}$ of having more spots on the second die than on the first, and $\mathbb{D}_{>} = \mathbb{D}_{\neq} \setminus \mathbb{D}_{<}$. $\mathbb{D}_{>}$ and $\mathbb{D}_{<}$ have the same cardinality because of the bijection $dd' \mapsto d'd$. Hence,

$$\mathrm{Pr}(\mathbb{D}_{>}) = \mathrm{Pr}(\mathbb{D}_{<}) = \frac{\mathrm{Pr}\left(\mathbb{D}_{\neq}\right)}{2} = \frac{5}{12} \;.$$

Flipping coins. Another classical probability space is that of a coin falling on one of its two sides, which results in head or tail with some given probability. Let us use $C = \{\textcircled{\scriptsize H}, \textcircled{\scriptsize T}\}$ for the set of elementary events, and let $\mathrm{Pr}(\textcircled{\scriptsize H}) = p$ and $\mathrm{Pr}(\textcircled{\scriptsize T}) = 1 - p$ for some $p \in \mathbf{R}$, $0 < p < 1$. If $p = \frac{1}{2}$, then we call the coin fair; otherwise, it is called biased. What if we want to model the experiment of repeatedly flipping a coin until we end up seeing head for the first time? Then

$$C' = \{\underbrace{\textcircled{\scriptsize H}}_{e_0}, \underbrace{\textcircled{\scriptsize T}\textcircled{\scriptsize H}}_{e_1}, \underbrace{\textcircled{\scriptsize T}\textcircled{\scriptsize T}\textcircled{\scriptsize H}}_{e_2}, \ldots\} \cup \{\underbrace{\textcircled{\scriptsize T}\textcircled{\scriptsize T}\textcircled{\scriptsize T}\cdots}_{e_\infty}\}$$

where we introduced some convenient shorthands for the elementary events. Here $\mathrm{Pr}(e_i) =$ $p(1-p)^i$, for $i \in \mathbf{N}_0$, and $\mathrm{Pr}(e_\infty) = 0$. At this point accept this as a definition and check that indeed $\sum_{i=0}^{\infty} p(1-p)^i = 1$. Let $C_0' = \{e_i \mid i \text{ even}\}$, the event of waiting an even number of tails until we succeed to see a head. Let $C_1' = \{e_i \mid i \text{ odd}\}$. We have

$$\mathrm{Pr}(C_0') = \sum_{i=0}^{\infty} p(1-p)^{2i} = p \sum_{i=0}^{\infty} \left((1-p)^2\right)^i = \frac{p}{1-(1-p)^2} = \frac{1}{2-p} \;.$$

Since C'_1 is – apart from a zero probability elementary event – the event complementary to C'_0, we have

$$\Pr(C'_1) = 1 - \frac{1}{2-p} = \frac{1-p}{2-p}\,.$$

Definition 2.2 (Random variable) *Given a probability space* $(\Omega, \Pr)$*, a* random variable *is a real-valued function*[a] *defined on the elementary events of a probability space, i.e.*

$$X : \Omega \to \mathbf{R}\,.$$

If A *is an event, then*

$$\omega \mapsto [\omega \in A]$$

is the indicator variable *for event* A*.*

[a] The fact that a random variable has to be real-valued is a restriction we apply here. In general, the image of a random variable may be just any set.

For the probability space of a single rolling die, we could consider the random variable X that maps the top side to the number of spots we see on this side

$$X : \boxed{1} \mapsto 1,\ \boxed{2} \mapsto 2,\ \boxed{3} \mapsto 3,\ \boxed{4} \mapsto 4,\ \boxed{5} \mapsto 5,\ \boxed{6} \mapsto 6\,,$$

or we could map to the number of spots on the invisible side sitting on the table

$$X' : \boxed{1} \mapsto 6,\ \boxed{2} \mapsto 5,\ \boxed{3} \mapsto 4,\ \boxed{4} \mapsto 3,\ \boxed{5} \mapsto 2,\ \boxed{6} \mapsto 1\,.$$

The mapping

$$Y : \boxed{1} \mapsto 0,\ \boxed{2} \mapsto 1,\ \boxed{3} \mapsto 0,\ \boxed{4} \mapsto 1,\ \boxed{5} \mapsto 0,\ \boxed{6} \mapsto 1$$

is the indicator variable for the event of seeing an even number of spots (previously denoted by D_{even}).

Note that X and X' depend on each other in the following sense. Suppose we know $X(d)$ for some $d \in D$, without knowing d itself, then we know also $X'(d)$, because $X(d) + X'(d) = 7$ for all $d \in D$. We write this as

$$X + X' = 7$$

for short, omitting the '(d)'. Similarly, X depends on Y, although not as explicitly as X depends on X'. Namely, $Y = 1$ tells us something about X: $Y(d) = 1 \Rightarrow X(d) \in \{2, 4, 6\}$, which we abbreviate as

$$Y = 1 \Rightarrow X \in \{2, 4, 6\}\,.$$

In contrast to this consider two random variables X_1 and X_2 on the space of two rolling dice. X_1 maps to the number of spots on the first die and X_2 to the number of spots on the second die. If $X_1(\omega) = 3$, say, we cannot give a better prediction for the value of $X_2(\omega)$. The same is true for any value possibly attained by X_1. So X_1 and X_2 have completely independent behavior.

For a last example in this context, let Z be the indicator variable for the event that the number of spots on the first die is at most the number of spots on the second die; we could write this as $Z := [X_1 \leq X_2]$. Now, $Z(\omega) = 1$ for some $\omega \in \mathbb{D}$ still allows all possible outcomes of $X_2(\omega)$. However, X_2 depends on Z in the sense that $Z = 1$ makes it more (and most) likely that $X_2 = 6$. That is, knowledge of Z allows a better prediction of X_2 (again, without seeing the underlying event).

Next we will formally capture this intuitive notion of independence.

Definition 2.3 (Independence) *Let* X *and* Y *be random variables defined on a common probability space. We say that* X *and* Y *are* independent random variables *if*

$$\Pr(X = x \wedge Y = y) \;=\; \Pr(X = x) \cdot \Pr(Y = y)$$

for all $x, y \in \mathbf{R}$*. A collection* X_i*,* $1 \leq i \leq n$ *of random variables on a common probability space is called* mutually independent *if*

$$\begin{aligned}\Pr(X_{i_1} = x_{i_1} \wedge X_{i_2} = x_{i_2} \wedge \cdots \wedge X_{i_k} = x_{i_k}) \quad &= \\ \Pr(X_{i_1} = x_{i_1}) \cdot \Pr(X_{i_2} = x_{i_2}) \cdot \cdots \cdot \Pr(X_{i_k} = x_{i_k})&\end{aligned}$$

for all $k \in \{2..n\}$*, all* $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ *and all* $(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) \in \mathbf{R}^k$*.*

Again, we have used some jargon: '$X = x$' short for the event $\{\omega \in \Omega \mid X(\omega) = x\}$, '$X = x \wedge Y = y$' short for $\{\omega \in \Omega \mid X(\omega) = x \wedge Y(\omega) = y\}$ etc.

It is important to realize that mutual independence is different from pairwise independence. For an example, consider the probability space of a pair of fair coins,

$$C^2 = \{ⓉⓉ, ⓉⒽ, ⒽⓉ, ⒽⒽ\}$$

with uniform distribution. Now we define three indicator variables

$$\begin{aligned} H_1 &:= [\text{first coin shows Ⓗ}]\,, \\ H_2 &:= [\text{second coin shows Ⓗ}]\,, \text{ and} \\ H_3 &:= [\text{exactly one coin shows Ⓗ}]\,. \end{aligned}$$

All three variables attain both 0 and 1 with probability $\frac{1}{2}$. We can verify that the variables are pairwise independent, but

$$\Pr(\text{first coin shows Ⓗ} \wedge \text{second coin shows Ⓗ} \wedge \text{exactly one coin shows Ⓗ}) = 0$$

and not $\frac{1}{8}$ as required for mutual independence.

Definition 2.4 (Expectation) *Let* X *be a random real-valued variable. The* expectation *(*expected value, mean*) of* X *is defined as*

$$\mathrm{E}(X) := \sum_{x \in X(\Omega+)} x \cdot \Pr(X = x) \tag{1}$$

provided this infinite sum exists.

For the example of rolling dice we have

$$\mathrm{E}(X) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5 \tag{2}$$

or for the random variable X^2

$$\mathrm{E}\big(X^2\big) = \sum_{i=1}^{6} i^2\frac{1}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6} = 15.16_{\ldots}$$

Note that this was just another shorthand. We used X^2 for the random variable

$$\omega \mapsto (X(\omega))^2 \; .$$

Observe that in our example $(\,\mathrm{E}(X))^2 = \frac{49}{4} \neq \frac{91}{6} = \mathrm{E}(X^2)$. Also, if X and Y are random variables, then we cannot expect $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$. Here XY stands for the random variable $\omega \mapsto X(\omega)Y(\omega)$.

Consider X and X' as defined for a single die. Then

$$\mathrm{E}(XX') = \frac{1\times 6 + 2\times 5 + 3\times 4 + 4\times 3 + 5\times 2 + 6\times 1}{6} = \frac{28}{3} = 9.33_{\ldots}$$

which is obviously not equal to $\mathrm{E}(X)\,\mathrm{E}(X') = \frac{49}{4} = 12.25$.

However, when it comes to linear functions of random variables, we have the following lemma, which is absolutely central for our investigations!

Lemma 2.1 (Linearity of Expectation) *Let X and Y be random variables defined on a common probability space and let $c \in \mathbf{R}$. Then*

$$\mathrm{E}(cX) = c\,\mathrm{E}(X) \;\; \textit{and} \;\; \mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)\; ,$$

provided $\mathrm{E}(X)$ and $\mathrm{E}(Y)$ exist.

Here is one typical route along which we will use the linearity of expectation. Recall the experiment of repeatedly flipping a coin until we see head for the first time, assuming that the coin flips are independent and head appears with probability p, $0 < p < 1$. What is the expectation for the number, X, of tails we see? Denote by X_i, $i \in \mathbf{N}$, the indicator variable that we see a tail in the ith round without having seen a head before. We have $X = \sum_{i=1}^{\infty} X_i$. $\Pr(X_i = 1) = (1-p)^i$, since this is a conjunction of i independent trials with success probability $(1-p)$. Hence,

$$\mathrm{E}(X_i) = 1 \times \Pr(X_i = 1) + 0 \times \Pr(X_i = 0) = \Pr(X_i = 1) = (1-p)^i\; ,$$

and[1]

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \mathrm{E}(X_i) = \sum_{i=1}^{\infty}(1-p)^i = \frac{1}{1-(1-p)} - 1 = \frac{1-p}{p}\; .$$

[1] Here we have to say "provided all expectations and sums involved exist!"

Note that the X_i's are not independent: $X_j = 1 \Rightarrow X_i = 1$ for all $i < j$, and so $\Pr(X_i = 1 \wedge X_j = 1) = (1-p)^j \neq (1-p)^{i+j}$.

I know that there is a direct way of deriving this expectation, since we know the probabilities $\Pr(X = i)$, $i \in \mathbf{N}_0$. But in many instances we will appreciate that it is possible to determine the expectation without knowing the distribution.

The lemma on the linearity of expectation made no request for independence, while this is required for a similar statement about the product of random variables.

Lemma 2.2 (Product of Independent Random Variables) *Let X and Y be two* independent *random variables defined on a common probability space. Then*

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)\; ,$$

provided $\mathrm{E}(X)$ and $\mathrm{E}(Y)$ exist.

Here is a small 'triviality' which is the core of the so-called probabilistic method, where one proves the existence of certain objects by analyzing random objects.

Lemma 2.3 (Existence from Expectation) *Let X be a random variable on a probability space $(\Omega, \Pr)$ for which the expectation $\mathrm{E}(X)$ exists. Then there exist elementary events ω_1 and ω_2 with*

$$X(\omega_1) \leq \mathrm{E}(X) \;\; \textit{and} \;\; X(\omega_2) \geq \mathrm{E}(X)\; .$$

Here is another simple fact which we will employ for deriving estimates for the probability that a random variable exceeds a certain value – so-called tail estimates.

Lemma 2.4 (Markov's Inequality) *Let X be a nonnegative random variable (i.e. $X(\Omega) \subseteq \mathbf{R}_0^+$) for which $\mathrm{E}(X)$ exists. Then, for all $\lambda \in \mathbf{R}^+$,*

$$\Pr(X \geq \lambda\,\mathrm{E}(X)) \leq \frac{1}{\lambda}.$$

Equality holds iff $X(\Omega^+) \subseteq \{0, \lambda\,\mathrm{E}(X)\}$.

Proof Let $t \in \mathbf{R}^+$.

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x \in X(\Omega^+)} x \cdot \Pr(X = x)\\
&= \sum_{x \in X(\Omega^+),\, x<t} \underbrace{x}_{\geq 0} \cdot \Pr(X = x) + \sum_{x \in X(\Omega^+),\, x \geq t} \underbrace{x}_{\geq t} \cdot \Pr(X = x)\\
&\geq t \cdot \sum_{x \in X(\Omega^+),\, x \geq t} \Pr(X = x)\\
&= t \cdot \Pr(X \geq t)
\end{aligned}$$

That is,

$$\Pr(X \geq t) \leq \frac{\mathrm{E}(X)}{t}\;, \;\; \text{for all } t \in \mathbf{R}^+.$$

Moreover, the inequality is strict iff there exists an $x \in X(\Omega^+)$ with $0 < x < t$, or there exists an $x \in X(\Omega^+)$ with $x > t$. It follows that both inequalities are identities iff $x \in X(\Omega^+)$ implies $x \in \{0, t\}$.

Now set $t = \lambda \operatorname{E}(X)$ to conclude the statement of the lemma. $\square$

We close this section with a short discussion of conditional probabilities and expectations.

Suppose somebody, call him Mr. McChance, offers you the following deal. First, you get 3.5 swiss francs. Then you have to roll two dice. If the second die shows a larger number of spots than the first one, you have to return that number (of spots on the second) of francs to friendly Mr. McChance; otherwise you have to return 2.5 francs to him. That may look quite attractive, at first glance, since the expected number of spots on the top face of the second rolling die is 3.5. And we even have some chance of paying 2.5 only. But then you play the game several times, and it looks like you are loosing. You are getting worried, and decide upon a thorough investigation of the game.

In order to analyze our expected gain or loss in the game, we have to distinguish two cases: The event $\mathbb{D}_{\geq}$ of the first die showing at least as many spots as the second, and the complementary event

$$\begin{aligned}
\mathbb{D}_{<} \;=\; \{ \;& \boxed{1}\boxed{2}, \boxed{1}\boxed{3}, \boxed{1}\boxed{4}, \boxed{1}\boxed{5}, \boxed{1}\boxed{6}, \\
& \boxed{2}\boxed{3}, \boxed{2}\boxed{4}, \boxed{2}\boxed{5}, \boxed{2}\boxed{6}, \\
& \boxed{3}\boxed{4}, \boxed{3}\boxed{5}, \boxed{3}\boxed{6}, \\
& \boxed{4}\boxed{5}, \boxed{4}\boxed{6}, \\
& \boxed{5}\boxed{6} \;\; \}
\end{aligned}$$

Here, of course, we see the pitfall of the procedure. *Given the event, that we have to pay the number of spots on the second die,* this number tends to be large – there is no configuration for that number to be 1, one for it to be 2, . . . , while there are 5 for it to be 6. Within the space of the 15 possible elementary events in $\mathbb{D}_{<}$, assuming uniform distribution among them, we expect to pay

$$2 \times \frac{1}{15} + 3 \times \frac{2}{15} + 4 \times \frac{3}{15} + 5 \times \frac{4}{15} + 6 \times \frac{5}{15} = \frac{70}{15} = 4.66\ldots \,.$$

So much for the bad news. But we may be lucky, $\mathbb{D}_{\geq}$ occurs (the chance for this to happen is 21 in 36 cases), and we have to pay 2.5 francs. We weight the two cases according to their probabilities to occur, and conclude that the expected number of francs we have to pay back is

$$\frac{70}{15} \times \frac{15}{36} + \frac{5}{2} \times \frac{21}{36} = \frac{245}{72} = 3.402\ldots \,.$$

So, after all, we have an expected gain of roughly 0.1 Swiss francs in the game. We can conclude that either (i) we made a mistake in our calculation, (ii) Mr. McChance brought loaded dice with him, (iii) bad luck, (iv) etc.

The analysis we just performed employs conditional probabilities – an essential tool in the analysis of randomized algorithms.

Definition 2.5 (Conditional Probabilities) *Let A and B be events in a probability space with $\Pr(B) > 0$. The* conditional probability of A, given B, *is defined to be*

$$\Pr(A \,|\, B) := \frac{\Pr(A \cap B)}{\Pr(B)} \,.$$

(In particular, let X and Y be random variables defined on a common probability space. Then the conditional probability of the event $X = x$, given the event $Y = y$, is

$$\Pr(X = x \,|\, Y = y) = \frac{\Pr(X = x \wedge Y = y)}{\Pr(Y = y)}$$

for all $x, y \in \mathbf{R}$ with $\Pr(Y = y) > 0$.)
Let X be a random variable and B be an event in a common probability space, $\Pr(B) > 0$. Then $X \,|\, B$ is the random variable obtained as the restriction of X to the probability space $(B, \Pr')$ with

$$\Pr' : \;\; \omega \mapsto \frac{\Pr(\omega)}{\Pr(B)} \,.$$

Conditional probabilities usually create some ‘notational confusion’, and let me just add to this by asking you to verify the identity

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

Anyways, a random variable has an expectation, and so has $X \,|\, Y = y$.

$$\begin{aligned}
\operatorname{E}(X \,|\, Y = y) &= \sum_{x \in (X|Y=y)((Y=y)^+)} x \Pr((X \,|\, Y = y) = x) \\
&= \sum_{x \in X(\Omega^+)} x \Pr(X = x \,|\, Y = y) \,.
\end{aligned}$$

provided the sum exists.

In the analysis of the game with McChance, we analyzed the random variable X for the amount we have to pay back after the experiment. Let Y be the indicator variable for the event $\mathbb{D}_{\geq}$. Then $\operatorname{E}(X \,|\, Y = 1) = 2.5$ and we calculated $\operatorname{E}(X \,|\, Y = 0) = 4.66\ldots$. The justification for our final step in the derivation of $\operatorname{E}(X)$ is given in the lemma below.

Lemma 2.5 (Laws of Total Probability and Expectation) *Let X and Y be two random variables on a common probability space $(\Omega, \Pr)$. Then*

$$\Pr(X = x) = \sum_{y \in Y(\Omega^+)} \Pr(X = x \,|\, Y = y) \Pr(Y = y) \tag{3}$$

for $x \in \mathbf{R}$, and

$$\operatorname{E}(X) = \sum_{y \in Y(\Omega^+)} \operatorname{E}(X \,|\, Y = y) \Pr(Y = y) \,, \tag{4}$$

provided $\operatorname{E}(X)$ exists.

Proof

$$\begin{aligned}
&\sum_{y\in Y(\Omega^+)} \Pr(X=x\,|\,Y=y)\Pr(Y=y)\\
&= \sum_{y\in Y(\Omega^+)} \frac{\Pr(X=x\wedge Y=y)}{\Pr(Y=y)}\Pr(Y=y)\\
&= \sum_{y\in Y(\Omega^+)} \Pr(X=x\wedge Y=y)\\
&= \Pr(X=x)
\end{aligned}$$

since every elementary event $\omega\in\Omega^+$ with $X(\omega)=x$ is mapped by Y to a unique $y\in Y(\Omega^+)$.

$$\begin{aligned}
&\sum_{y\in Y(\Omega^+)} \mathrm{E}(X\,|\,Y=y)\Pr(Y=y)\\
&= \sum_{y\in Y(\Omega^+)} \left(\sum_{x\in X(\Omega^+)} x\Pr(X=x\,|\,Y=y)\right)\Pr(Y=y)\\
&= \sum_{x\in X(\Omega^+)} x \sum_{y\in Y(\Omega^+)} \Pr(X=x\,|\,Y=y)\Pr(Y=y)\\
&= \sum_{x\in X(\Omega^+)} x\Pr(X=x)\\
&= \mathrm{E}(X)
\end{aligned}$$

$\square$

In the game with McChance, we derived

$$\mathrm{E}(X) = \underbrace{\mathrm{E}(X\,|\,Y=0)}_{70/15}\underbrace{\Pr(Y=0)}_{15/36} + \underbrace{\mathrm{E}(X\,|\,Y=1)}_{5/2}\underbrace{\Pr(Y=1)}_{21/36} .$$

Exercise 2.1 *Let X be a nonnegative random variable and $s,t\in\mathbf{R}_0^+$, $s<t$. What can you say about $\mathrm{E}(X)$ in terms of $\Pr(X<s)$ and $\Pr(X\geq t)$?*

Exercise 2.2 *Let $X_1,X_2,\ldots,X_n$ be $\{0,1\}$-valued random variables defined on a common probability space. Show that they are mutually independent iff*

$$\Pr\left(\bigwedge_{i\in I}(X_i=1)\right) = \prod_{i\in I}\Pr(X_i=1)$$

for all $I\subseteq\{1..n\}$.

Exercise 2.3 Poisson Distribution

$\lambda\in\mathbf{R}^+$. *An $\mathbf{N}_0$-valued random variable X is said to have the* Poisson distribution with parameter λ, *if $\Pr(X=i)=\frac{\lambda^i}{i!}e^{-\lambda}$ for all $i\in\mathbf{N}_0$. Show that this is indeed a probability distribution and determine its expectation. Moreover, show that if X and Y have the Poisson distribution with parameters λ and μ, respectively, then $X+Y$ is a random variable with Poisson distribution.*

3 Left-to-Right Minima in a Permutation

$n\in\mathbf{N}$. Consider a random permutation $(a_1,a_2,\ldots,a_n)$ from the uniform distribution on $\mathcal{S}_n$. We read the permutation from left to right, and we count how often the minimum of the numbers seen so far changes. That is, we investigate the random variable

$$X := \#\{i\in\{1..n\}\mid a_i<\min\{a_1,a_2,\ldots,a_{i-1}\}\} .$$

X, or related distributions, occur frequently in algorithms. But, perhaps even more important, we will encounter (or recall) some simple methods.

Backwards analysis. In order to analyze the expectation of X, we let X_i, $1\leq i\leq n$, be the indicator variable for the event that position i holds a left-to-right minimum;

$$X_i := [a_i<\min\{a_1,a_2,\ldots,a_{i-1}\}] .$$

Clearly, $X=\sum_{i=1}^n X_i$. So knowing the expectations of the X_i's, or equivalently, the probabilities

$$\Pr(a_i<\min\{a_1,a_2,\ldots,a_{i-1}\}) = \Pr(a_i=\min\{a_1,a_2,\ldots,a_i\})$$

will suffice.

Let us generate the random permutation from left to right, i.e. we choose a_i uniformly at random from $\{1..n\}\setminus\{a_1,a_2,\ldots,a_{i-1}\}$, for $i=1,2,\ldots,n$. (Check here that indeed every permutation is generated with probability $\frac{1}{n!}$ in this way.) It seems difficult to analyze what the probability of $a_i=\min\{a_1,a_2,\ldots,a_i\}$ is; after all, that heavily depends on the numbers chosen so far. Here comes a simple twist to our assistance: Although the problem statement suggests that we generate the permutation from left to right, we should do it in the other direction, *'backwards'*, so to say. That is, we choose a_i uniformly at random from $\{1..n\}\setminus\{a_{i+1},a_{i+2},\ldots,a_n\}$, for $i=n,n-1,\ldots,1$. Now, independently of the choices in $\{a_{i+1},a_{i+2},\ldots,a_n\}$,

$$\Pr(a_i=\min(\{1..n\}\setminus\{a_{i+1},a_{i+2},\ldots,a_n\})) = \frac{1}{i} .$$

We can conclude that

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{E}(X_i) = \sum_{i=1}^n \Pr(X_i=1) = \sum_{i=1}^n \frac{1}{i} = H_n .$$

The nth Harmonic number. More than once we will encounter H_n, the nth Harmonic number, in our analyses. So it is worthwhile to know some of its properties. H_n looks like a discrete analogue of $\int_1^n \frac{dt}{t}=\ln n$. In fact, an appropriate picture of the graph of the function $\frac{1}{t}$ will show you that

$$\ln(n+1)\leq H_n\leq 1+\ln n\,,\quad \text{for } n\geq 1.$$

$H_n-\ln n$ converges to a constant γ as n goes to infinity. $\gamma=0.57721\ldots$ is called Euler's constant, and – as far as I know – it is not known whether this number is rational or irrational.

Higher moments. Our backwards analysis shows that the X_i's are pairwise independent. Hence, we can calculate $\mathrm{E}(X^2) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}(X_i X_j)$. On one hand, for $i \neq j$, $\mathrm{E}(X_i X_j) = \mathrm{E}(X_i)\,\mathrm{E}(X_j) = \frac{1}{ij}$. On the other hand, we have $\mathrm{E}(X_i^2) = \frac{1}{i} = \frac{1}{i^2} + (\frac{1}{i} - \frac{1}{i^2})$. Hence,

$$\mathrm{E}\Big(X^2\Big) = (H_n)^2 + H_n - \sum_{i=1}^n \frac{1}{i^2} \; .$$

(The sum $\sum_{i=1}^\infty \frac{1}{i^2}$ converges to[2] $\zeta(2) = \frac{\pi^2}{6}$.) Therefore,

$$\mathrm{var}(X) := \mathrm{E}\Big((X - \mathrm{E}(X))^2\Big) = \mathrm{E}\Big(X^2\Big) - \mathrm{E}(X)^2 = H_n - \sum_{i=1}^n \frac{1}{i^2} \; .$$

Now, employing Markov's Inequality, we may conclude that

$$\begin{aligned} \Pr\Big(|X - \mathrm{E}(X)| \geq \lambda\sqrt{H_n}\Big) &\leq \Pr\Big(|X - \mathrm{E}(X)| \geq \lambda\sqrt{\mathrm{var}(X)}\Big) \\ &= \Pr\Big((X - \mathrm{E}(X))^2 \geq \lambda^2\,\mathrm{var}(X)\Big) \\ &\leq \frac{1}{\lambda^2} \end{aligned}$$

for $\lambda \in \mathbf{R}^+$, since $\sqrt{H_n} > \sqrt{\mathrm{var}(X)}$. This application of Markov's Inequality to $(X - \mathrm{E}(X))^2$ via the variance is called *Chebyshev's Inequality*.

Double check. It is always worthwhile to check a result on small numbers or examples. The fact that we have exact results instead of asymptotic bounds, allows us to do so. So here are the permutations in $\mathcal{S}_3$ with left-to-right minima underlined, and the numbers of such minima.

permutation	X
$\underline{1}\;2\;3$	1
$\underline{1}\;3\;2$	1
$\underline{2}\;\underline{1}\;3$	2
$\underline{2}\;3\;\underline{1}$	2
$\underline{3}\;\underline{1}\;2$	2
$\underline{3}\;\underline{2}\;\underline{1}$	3

So for $n = 3$, we have $\mathrm{E}(X) = \frac{2\times1+3\times2+1\times3}{6} = \frac{11}{6}$ which, indeed, equals $1 + \frac{1}{2} + \frac{1}{3} = H_3$. Moreover, $\mathrm{E}(X^2) = \frac{2\times1+3\times4+1\times9}{6} = \frac{23}{6}$, which we have to compare with

$$\left(\frac{11}{6}\right)^2 + \frac{11}{6} - \left(1 + \frac{1}{4} + \frac{1}{9}\right) = \frac{121 + 66 - 36 - 9 - 4}{36} = \frac{138}{36} = \frac{23}{6} \; .$$

Just to avoid any misunderstanding: Such checking of small examples (numbers) does not add *anything* to the correctness of a result or a proof, but it may *falsify* a result – which is extremely helpful. In this spirit, a good proof should be written in a way, that it is indeed easy to falsify, if it is not correct!

[2] ζ the Riemann zeta function.

Analysis by counting. Perhaps you are not convinced by the backwards argument above So let us do the proof by brute force counting. Let $1 \leq i \leq n$. How many permutations have a left-to-right minimum at position i? We can generate such a permutation by first choosing any $n - i$ numbers in $\{1..n\}$ for positions $i + 1$ to n, and place them in these positions in $(n - i)!$ ways; gives $\binom{n}{n-i}(n - i)!$. Then we have to choose the smallest among the remaining numbers for position i; gives one option only. Finally, we can place the remaining $i - 1$ numbers in $(i - 1)!$ ways in positions 1 to $i - 1$; gives us $(i - 1)!$ possibilities. We have counted

$$\binom{n}{n-i}(n - i)! \times 1 \times (i - 1)! = \frac{n!}{i}.$$

There are $n!$ permutations altogether, so if we choose a random one uniformly among those, with probability $\frac{n!/i}{n!} = \frac{1}{i}$ position i will contain a left-to-right minimum.

Given $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, we can calculate that

$$\Pr\left(X_{i_1} = 1 \wedge X_{i_2} = 1 \wedge \ldots \wedge X_{i_k} = 1\right) = \frac{1}{i_1 i_2 \cdots i_k}$$

which shows mutual independence of the variables X_i (consult exercise in previous chapter). This gives us access to the expectation $\mathrm{E}\Big(2^X\Big)$.

Even higher moments. Note that[3]

$$2^X = \sum_{A \in 2^{\{1..n\}}} [X_i = 1 \text{ for all } i \in A] \; .$$

You may have to think about this. After clarifying the identity, you may be annoyed, since we managed to write something simple in a complicated way. However, we know that

$$\Pr\left(X_i = 1 \text{ for all } i \in A\right) = \frac{1}{\prod_{i \in A} i}$$

and, due to the blessing of linearity of expectation, we conclude for $n > 1$

$$\mathrm{E}\Big(2^X\Big) = \sum_{A \in 2^{\{1..n\}}} \frac{1}{\prod_{i \in A} i} = \sum_{A \in 2^{\{1..n-1\}}} \frac{1}{\prod_{i \in A} i} + \left(\frac{1}{n} \cdot \sum_{A \in 2^{\{1..n-1\}}} \frac{1}{\prod_{i \in A} i}\right)$$

That is, if we set $f_n = \mathrm{E}\Big(2^X\Big)$ for permutations of length n, then $f_n = f_{n-1} + \frac{1}{n} f_{n-1}$ for $n > 1$, and $f_1 = 2$. We can easily verify that $f_n = n + 1$. Following our previous advice we look at $\mathcal{S}_3$ and calculate

$$\mathrm{E}\Big(2^X\Big) = \frac{2 \times 2^1 + 3 \times 2^2 + 1 \times 2^3}{6} = \frac{24}{6} = 4 \; .$$

[3] Recall that $[S]$ is the indicator function for a statement S that can be true or false.

And, again, we get a tail estimate by applying Markov's Inequality, now to 2^X.

$$\begin{aligned} \Pr(X \ge (1+\lambda)\lg(n+1)) &= \Pr\left(2^X \ge 2^{(1+\lambda)\lg(n+1)}\right) \\ &= \Pr\left(2^X \ge (n+1)^{1+\lambda}\right) \\ &= \Pr\left(2^X \ge (n+1)^{\lambda}\,\mathrm{E}\left(2^X\right)\right) \\ &\le (n+1)^{-\lambda} \end{aligned}$$

for all $\lambda \in \mathbf{R}^+$.

I have to admit that there is a more direct way of obtaining $\mathrm{E}\left(2^X\right)$. Namely,

$$\mathrm{E}\left(2^{X_i}\right) = 2^1 \times \frac{1}{i} + 2^0 \times \left(1 - \frac{1}{i}\right) = \frac{i+1}{i}$$

which leads to

$$\mathrm{E}\left(2^X\right) = \mathrm{E}\left(2^{\sum_{i=1}^n X_i}\right) = \mathrm{E}\left(\prod_{i=1}^n 2^{X_i}\right) = \prod_{i=1}^n \mathrm{E}\left(2^{X_i}\right) = \prod_{i=1}^n \frac{i+1}{i} = n+1\,.$$

Note that here it was important that the X_i's are mutually independent! Now, of course, this derivation raises the question why we should restrict ourselves to a base of 2. Recall that we derived a tail estimate via this exponential moment, and perhaps a base different from 2 allows a better estimate.

Yet another analysis of $\mathrm{E}(X)$. What is the probability that the number n is a left-to-right minimum in the random permutation $(a_1, a_2, \ldots, a_n)$? This is the case iff $a_1 = n$, which happens with probability $\frac{1}{n}$. Adding n to a permutation of $\{1..n-1\}$ will not change the number of left-to-right minima, unless n is first, when the number of such minima is increased by 1. So if we denote by f_n the expected number of left-to-right minima in a random permutation in $\mathcal{S}_n$, $n \in \mathbf{N}$, then $f_1 = 1$ and $f_n = f_{n-1} + \frac{1}{n}$, which immediately shows $f_n = H_n$. A closer look at the analysis reveals that we obtained the result here on a path different from our previous analysis. Namely, we looked at indicator variables Y_i for the event that the number i occurs as left-to-right minimum. It is easy to derive $\mathrm{E}(Y_i) = \frac{1}{i}$, since $Y_i = 1$ iff i is first among $\{1..i\}$.

A game. Here is a randomized game, or a stochastic process, if you like. You start with a number $n \in \mathbf{N}$. If $n = 1$, nothing happens. If $n > 1$, a number i is chosen uniformly at random from $\{1..n-1\}$, and we continue with this number i as with n before. What is the expected number of steps it takes to finish the process?

Let Z_n, $n \in \mathbf{N}$, be the random variable for the number of steps when the process is started with the number n. $Z_1 = 0$. If $n > 1$, we partition the games according to the first number chosen – this invokes conditional expectations. Namely, let F be the random variable for this first number chosen, then

$$\begin{aligned} \mathrm{E}(Z_n) = \quad & \mathrm{E}(Z_n \,|\, F = 1) \cdot \Pr(F = 1) \\ + \quad & \mathrm{E}(Z_n \,|\, F = 2) \cdot \Pr(F = 2) \\ + \quad & \ldots \\ + \quad & \mathrm{E}(Z_n \,|\, F = n-1) \cdot \Pr(F = n-1) \end{aligned}$$

Since

$$\mathrm{E}(Z_n \,|\, F = i) = 1 + \mathrm{E}(Z_i) \quad \text{and} \quad \Pr(F = i) = \frac{1}{n-1}, \quad 1 \le i \le n-1,$$

we have

$$\mathrm{E}(Z_n) = \sum_{i=1}^{n-1} (1 + \mathrm{E}(Z_i)) \frac{1}{n-1} = 1 + \frac{1}{n-1} \sum_{i=1}^{n-1} \mathrm{E}(Z_i)\ , \quad \text{for } n > 1.$$

Let us use z_n short for $\mathrm{E}(Z_n)$. Then $z_1 = 0$ and $z_n = 1 + \frac{1}{n-1} \sum_{i=1}^{n-1} z_i$, or equivalently

$$(n-1) z_n = (n-1) + (z_1 + z_2 + \ldots + z_{n-2} + z_{n-1}), \quad \text{for } n > 1.$$

Now we can subtract this identity with $n-1$ substituted for n

$$(n-2) z_{n-1} = (n-2) + (z_1 + z_2 + \ldots + z_{n-2}), \quad \text{for } n > 2.$$

which gives, for $n > 2$,

$$\begin{aligned} (n-1) z_n - (n-2) z_{n-1} &= 1 + z_{n-1} \\ (n-1) z_n &= 1 + (n-1) z_{n-1} \\ z_n &= \frac{1}{n-1} + z_{n-1} \end{aligned}$$

The Harmonic number is back again! Since $z_1 = 0$ and $z_2 = 1$, we have derived $z_n = H_{n-1}$, i.e.

$$\mathrm{E}(Z_n) = H_{n-1}, \quad \text{for } n \in \mathbf{N}.$$

New or old? We have seen that, for $n \ge 2$, $\mathrm{E}(Z_n)$ equals the expected number of left-to-right minima in a random permutation in S_{n-1}. This may be purely incidental; e.g. the underlying distributions may be completely different, so we cannot carry over the higher moments we have derived for left-to-right minima.

Here is how the two distributions relate to each other. Note that our game requires a random source that generates the random numbers as required. Suppose we had some access to random permutations, how could we use that for the game? At the beginning of the game starting with n we request a random permutation $(a_1, a_2, \ldots, a_{n-1})$ in S_{n-1}. Then, we let a_1 be the first random number needed in the game. Clearly, a_1 is a random number uniform in $\{1..n-1\}$. When we need the next number, i.e. a random number in $\{1..a_1\}$ (unless $a_1 = 1$), then we first try a_2, if that is not smaller than a_1, we try a_3 and so on. The next useful number is the next left-to-right minimum in the permutation! And if we proceed like this, the number of left-to-right minima will be the number of steps our game will take.

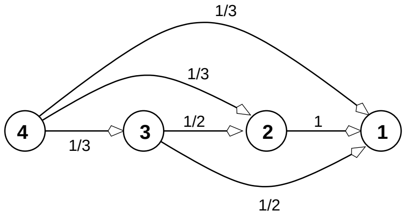
If we play the game as described above, we see that the two distributions – steps in the game and left-to-right minima – are the same. But do we really get the random numbers according to the right distribution? To this end, imagine that you generate the distribution on the fly from left to right by choosing a_i uniformly at random from $\{1..n\} \setminus \{a_1, a_2, \ldots, a_{i-1}\}$. The numbers exceeding the current left-to-right minimum will be ignored (for the game), and the first number smaller than the current left-to-right minimum will indeed be uniform among all such numbers.

So, after all, all we have derived for X can now be reused for Z_{n+1}.

Random paths in transitive tournaments. The transitive tournament T_n, $n \in \mathbf{N}$, is the directed graph with vertex set $V = V(T_n) = \{1..n\}$ and edge set

$$E = E(T_n) = \{(i,j) \in V^2 \mid n \geq i > j \geq 1\} .$$

Every sequence of numbers $n = i_0 > i_1 > \ldots > i_k = 1$ describes a possible path from n to 1 of length k. Suppose we follow such a path in a random manner by always choosing at a vertex an outgoing edge uniformly at random from all outgoing edges. It is easy to see that we generate the same distribution on paths as in the game considered above; it's just a different point of view.



Exercise 3.1 Harmony in the Exponent

Find good upper and lower bounds for e^{H_n}, $n \in \mathbf{N}$.

Exercise 3.2 Sum of Harmonic Numbers

What can you say about

$$\sum_{i=1}^{n} H_i \text{ and } \sum_{i=1}^{n} iH_i, \quad n \in \mathbf{N} \text{ ?}$$

Exercise 3.3 Almost Riemann Zeta

What can you say about

$$\sum_{i=1}^{n} \frac{1}{i(i+1)}, \quad n \in \mathbf{N} \text{ ?}$$

Exercise 3.4 A Product

What can you say about

$$\prod_{i=1}^{n} \left(1 + \frac{2}{i}\right), \quad n \in \mathbf{N} \text{ ?}$$

Exercise 3.5 Qualities of Inequalities

Consider the random variable X for the number of left-to-right minima in a random permutation in $\mathcal{S}_n$. For $n = 1000$, compare the tail estimates $\Pr(X \geq 50)$ we get (i) via Markov's Inequality applied to X, (ii) via Chebyshev's Inequality (i.e. Markov's applied to $(X - \mathrm{E}(X))^2$), and (iii) via Markov's Inequality applied to 2^X.

Exercise 3.6 Paranoia

Here is an algorithm for computing the minimum of n distinct numbers, $n \in \mathbf{N}$. First we permute the numbers at random (uniform). Then we perform the following steps on this permutation $(a_1, a_2, \ldots, a_n)$.

```
$m \leftarrow \infty$;
for $i \leftarrow 1$ to $n$ do
    if $a_i < m$ then
        $m \leftarrow a_i$;
        for $j \leftarrow 1$ to $i$ do
            if $a_j < m$ then output “Something wrong!”;
output “$m$ is the minimum”;
```

So the algorithm compares a next element with the previous minimum. If we have encountered a smaller element, we check it against all previous numbers – not trusting the transitivity of the $<$ relation. What is the expected number of comparisons (“$a_i < m$” or “$a_j < m$”) of the algorithm? What is the maximum and what is the minimum number of comparisons necessary?

Exercise 3.7 Left-to-right Minima and Maxima

$n \in \mathbf{N}$. *Consider a random permutation $(a_1, a_2, \ldots, a_n)$ of $\{1..n\}$. How often does the minimum or maximum change from left to right? More precisely, let*

$$M_i := (\min\{a_1, a_2, \ldots, a_i\}, \max\{a_1, a_2, \ldots, a_i\}) , \quad \textit{for } 0 \leq i \leq n.$$

What is the expected cardinality of $\{i \in \{1..n\} \mid M_i \neq M_{i-1}\}$?

Exercise 3.8 An Even Higher Moment

Consider X, the random variable for the number of left-to-right minima in a permutation uniform from $\mathcal{S}_n$. We have determined $\mathrm{E}\big(2^X\big)$. What is $\mathrm{E}\big(3^X\big)$? Use the expectation for a tail estimate. Is it better than the one obtained via $\mathrm{E}\big(2^X\big)$?

Exercise 3.9 A Low Moment

Consider X, the random variable for the number of left-to-right minima in a permutation uniform from $\mathcal{S}_n$. What is $\mathrm{E}\left(2^{-X}\right)$? Use the expectation for a tail estimate for $\Pr(X < t)$.

Exercise 3.10 Simultaneous Left-to-Right Minima

$n \in \mathbf{N}$. *Consider two random permutations $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$, independently drawn from the uniform distribution of $\mathcal{S}_n$. What is the expected number of indices i, such that both a_i and b_i are left-to-right minima?*

Exercise 3.11 Change of Smallest Distance

$n \in \mathbf{N}$. *We are given n points in the plane, all distances between the points distinct. Now we consider a random permutation $(p_1, p_2, \ldots, p_n)$ of the points and count how often the smallest distance between two points in $\{p_1, p_2, \ldots, p_i\}$ changes, for $i = 2, 3, \ldots, n$? What is the expected number of such changes? What can you say, if some of the pairs of points have the same distance?*

Exercise 3.12 Embracing Zero

You start with a pair (n, m) of natural numbers. If $n = m = 1$ you are done. Otherwise, choose a random number uniform in

$$\{i \in \mathbf{Z} \mid -n < i < m, i \neq 0\} .$$

If i is negative, move to the pair $(-i, m)$; if i is positive, move to (n, i). What is the expected number of steps until you are done?

Exercise 3.13 Survival of the First King

$n \in \mathbf{N}$. *Consider a random permutation $(a_1, a_2, \ldots, a_n)$ of $\{1..n\}$. What is the expectation for $\min\{i \in \{1..n\} \mid a_i < a_1\}$?*

Exercise 3.14 Selection of a Leader

$n \in \mathbf{N}$. *n people are sitting around a table, and each of them gets a random number in $\{1..n\}$, all distinct. Now everybody looks at her number i and checks the distance cw_i to the first number $x \leq i$ in clockwise direction; e.g. if the immediate clockwise neighbor has a smaller number, then $cw_i = 1$; if $i = 1$, then $cw_i = n$. What is the expectation for the sum $\sum_{i=1}^{n} cw_i$? What are upper and lower bounds for this sum?*

Exercise 3.15 Amongst Light Bulbs

$n \in \mathbf{N}$. *You are sitting on a circle with n light bulbs, all initially dark. Now these light bulbs are lit up, one at a time in random order. However, you realize such a new light bulb to lit up only if all light bulbs along the circle (in clockwise* or *counterclockwise direction) between you and this light bulb are still dark. So you definitely see the first and second, but there is a chance that you miss the third. What is the expected number of light bulbs you see being lit up?*

Exercise 3.16 A Light Bulb Amongst Light Bulbs

$n \in \mathbf{N}$. *There are n light bulbs on a circle, all initially dark, and you are one of them. Now these light bulbs are lit up, one at a time in random order. However, you realize such a new light bulb to lit up only if all light bulbs along the circle (in clockwise* or *counterclockwise direction) between you and this light bulb are still dark, and only as long you yourself are still dark. So you see the first one, unless its yourself. What is the expected number of light bulbs you see being lit up?*

Exercise 3.17 Accumulating Left-to-Right Minima

Given a random permutation of $\{1..n\}$, $n \in \mathbf{N}$, we add up the numbers which appear as left-to-right minima. What is the expected sum? What if you add up the left-to-right maxima?

Exercise 3.18 Probability of Getting All Primes

Suppose we consider the game described in this section, and we start the game with $n = 13$. What is the probability that the sequence of numbers chosen in the course of the game is $(11, 7, 5, 3, 2, 1)$?

Exercise 3.19 Somewhat Slower Game

Here is a slight variation to the game from this section. You start with a number $n \in \mathbf{N}$. If $n = 1$, nothing happens. If $n > 1$, a random number i uniform from $\{1..n\}$ is chosen, and we continue with this number i as with n before. So the difference is that we can now choose the number n as well! What is the expected number of steps it takes to finish the process?

Exercise 3.20 An Even Slower Game

You start with a natural number n, $n \geq 2$. If $n = 2$, nothing happens. If $n > 2$, two distinct random numbers $\{a, b\}$ uniform from $\binom{\{1..n\}}{2}$ are chosen, and we continue with the number $\max\{a, b\}$ as with n before. What is the expected number of steps it takes to finish the process?

Exercise 3.21 High Expectations?

Consider an $\mathbf{N}$-valued random variable with $\Pr(X = i) = \frac{1}{i(i+1)}$ for all $i \in \mathbf{N}$. Show that this is indeed a probability distribution, and determine the expectations of X, and of the random variables $Y := \frac{1}{X}$ and $Z := 1 + \frac{1}{X}$.

4 Independent Sets

An independent set in an undirected graph $G = (V, E)$, ($V = \{1..n\}$ and $E \subseteq \binom{V}{2}$), is a subset I of V with $E \cap \binom{I}{2} = \emptyset$. Finding the maximum cardinality of an independent set is NP-hard, and actually it is hard to approximate this number, unless $P = NP$.

Here we consider several ways of taking random independent sets. (m denotes the number of edges of G.)

Taking random subsets of a given size. How many vertices can we choose, so that we can expect less than one edge among those vertices? Fix some natural number $r < n$. If we choose a set R of r vertices at random (every set in $\binom{V}{r}$ with the same probability), then the probability for an edge $\{i, j\}$ to have both endpoints in R is

$$\frac{\binom{n-2}{r-2}}{\binom{n}{r}} = \frac{r(r-1)}{n(n-1)} .$$

For the random variable $X := \#(\binom{R}{2} \cap E)$ we compute

$$\mathrm{E}(X) = m \cdot \frac{r(r-1)}{n(n-1)} < \frac{mr^2}{n^2} .$$

This quantity is less than one if r is chosen to be at most $n/\sqrt{m}$. That is, there must exist an independent set of $\lfloor n/\sqrt{m} \rfloor$.

Now suppose you are actually trying to find a large independent set in a given graph. If we choose $r = \lfloor n/\sqrt{2m} \rfloor$, then $\mathrm{E}(X) < 1/2$. By Markov's Inequality,

$$\Pr(X \geq 1) \;\leq\; \Pr(X \geq 2\,\mathrm{E}(X)) \;\leq\; \frac{1}{2} .$$

Therefore

$$\Pr(X = 0) = \Pr(X < 1) = 1 - \Pr(X \geq 1) \geq 1 - \frac{1}{2} = \frac{1}{2} .$$

That is, a random set of $\lfloor n/\sqrt{2m} \rfloor$ vertices is an independent set with probability at least $\frac{1}{2}$.

Suppose you have a graph and you repeat the experiment several times with $r = 100$ and you don't get an independent set. Once you are close: there is only one edge e among your set M of 100 vertices chosen. What would you do? Of course, you wouldn't reject M completely. Instead, you can remove one of the endpoints of e and you have an independent set of size 99. We will see how this idea provides independent sets of much larger size in general!

Repairing small defects. Let us fix a real number p, $0 \leq p \leq 1$. We choose $S \subseteq V$ by taking every vertex into S with probability[4] p. For each edge in $F = \binom{S}{2} \cap E$, we remove one endpoint from S and we obtain an independent set of size at least $\#S - \#F$ in this way. Let X be the random variable for $\#S$ and let Y be the random variable for $\#F$. We obtain

$$\mathrm{E}(X - Y) = \mathrm{E}(X) - \mathrm{E}(Y) = np - mp^2 ,$$

and this expression is maximized by $p = n/(2m)$. For that value it yields $\mathrm{E}(X - Y) = n^2/(4m)$ (note that we assume here $m > n/2$; otherwise $p > 1$). This is much better than our previous bound of $\frac{n}{\sqrt{m}}$.

[4] This is similar to taking a random subset of size pn.

Random ordering. Let π be an ordering of the vertices of G. We define a set $S_\pi \subseteq V$, by letting $i \in S_\pi$ if and only if none of its neighbors precedes i in the order π. Clearly, S_π is an independent set in G. What is its expected size, if π is chosen uniformly at random from $\mathcal{S}_n$? For $i \in V$, let X_i be the indicator variable for the event that i is in S_π. If d_i denotes the number of neighbors of i in G, then the probability of i to lie in S_π is $1/(d_i + 1)$. It follows that

$$\mathrm{E}(\#S_\pi) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} \frac{1}{d_i + 1} . \tag{5}$$

This bound is probably not very transparent, so let us try to conclude a bound in terms of $m = \#E$. To this end, observe that[5]

$$\left(\frac{\sum_{i=1}^{n}(d_i+1)^{-1}}{n}\right)^{-1} \leq \frac{\sum_{i=1}^{n}(d_i+1)}{n} = \frac{2m+n}{n} ,$$

and so

$$\mathrm{E}(\#S_\pi) \geq \frac{n^2}{2m+n} .$$

If G consists of k cliques of size s, then $m = k\binom{s}{2}$, and $n = ks$. Hence, the bound gives

$$\frac{k^2 s^2}{ks(s-1) + ks} = k ,$$

which is obviously tight. (Indeed, bound (1) implies already Turán's Theorem, see [1, page 81].)

Note, that in an algorithmic setting, we could look at the vertices of G in the order given by π and add a vertex to a set T, if none of its neighbors are in T yet. In this way we get a set $T_\pi \supseteq S_\pi$, which may be much larger than S_π. For example, if G is a complete bipartite graph with independent sets of size s, $n = 2s$, then $\mathrm{E}(\#S_\pi) = n/(s+1) = 2n/(n+2) < 2$, while T_π will always contain $n/2$ vertices.

Exercise 4.1 A Quadratic Program

Let $G = (\{1..n\}, E)$ be a graph. Show that the solution to the quadratic program

$$\begin{array}{ll} \textit{maximize} & \displaystyle\sum_{i=1}^{n} p_i - \sum_{\{i,j\} \in E} p_i p_j \\ \textit{subject to} & 0 \leq p_i \leq 1 \quad \textit{for all} \quad 1 \leq i \leq n \end{array}$$

equals the maximum cardinality of an independent set.

Exercise 4.2 Large Cuts

Let $G = (V, E)$ be a graph. A cut is a partition of V into two sets, i.e. $C = \{S, V \setminus S\}$ for some $S \subseteq V$. The set of edges in the cut, $E(C)$, is defined by $\{\{x,y\} \in E \mid \#(\{x,y\} \cap S) = 1\}$. Show that there is always a cut C with $E(C) \geq \#E/2$. Show that, if $E \neq \emptyset$, then there is exists always a cut of size strictly larger than $\#E/2$.

[5] Harmonic mean cannot exceed arithmetic mean.

Exercise 4.3 Three-Colored Triangles
Let $G = (V, E)$ be a graph with t triangles. Show that it is always possible to color the vertices with three colors, such that $\frac{2t}{9}$ triangles have their vertices colored with three distinct colors. What can you say about a coloring with four colors (still, you want to count the triangles that have no two vertices colored with the same color)? What can you say about the number of triangles that are not monochromatic (not all vertices the same color)?

Exercise 4.4 Higher Moments for Random Cuts
A random cut $\{S, V \setminus S\}$ in a graph $G = (V, E)$ is chosen by letting each vertex to be in S with probability $1/2$, independently from the other vertices. Let X be the random variable for the size of such a random cut in a graph with n vertices and m edges. Determine $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$, and apply Chebyshev's Inequality for a tail estimate. What can you say about $\mathrm{E}(X^3)$ or 2^X?

Exercise 4.5 Points at Given Minimum Distance
Let S be a set of n points in the plane, no two at distance smaller than one from each other. Find a function $f(n)$ (as large as you can), such that such a set has always a subset S' of size $f(n)$ with no two points at distance smaller than two from each other.

Exercise 4.6 No or One Neighbor
$n \in \mathbf{N}$. We are given a graph $G = (\{1..n\}, E)$ and a permutation $\pi = (v_1, v_2, \ldots, v_n)$ in $\mathcal{S}_n$. Now consider the following procedure that generates a subset F_π of the vertices of G.

```
$F \leftarrow \emptyset$;
for $i \leftarrow 1$ to $n$ do
    if $v_i$ has no or one neighbor in $F$ then
        $F \leftarrow F \cup \{v_i\}$;
output $F$;
```

What can you say about the subgraph of G induced by F? Give estimates for the expected size of F_π, π a random permutation, in terms of n and $m = \#E$.

Exercise 4.7 Triangle-Free Induced Subgraphs with Many Vertices
Find some function $f(n, t)$ (as large as you can) such that every graph with n vertices and t triangles has a triangle-free induced subgraph with $f(n, t)$ vertices.

Exercise 4.8 Triangle-Free Subgraphs with Many Edges
Find some function $f(n, m, t)$ (as large as you can) such that every graph with n vertices, m edges, and t triangles has a triangle-free subgraph (not necessarily induced!) with $f(n, m, t)$ edges.

Exercise 4.9 Dominating Sets
A dominating set *of an undirected graph $G = (V, E)$ is a subset D of V such that every vertex $v \in V \setminus U$ has at least one neighbor in U. Propose probabilistic methods for generating small dominating sets. Analyze, what you can guarantee for r-regular graphs (all vertices have same degree r) with n vertices.*

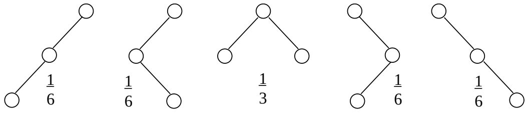
Exercise 4.10 Bob and Clarissa
Bob and Clarissa were challenged to find a graph with 10 vertices, many edges and only few cycles. Clarissa has the best example so far, but then Bob sends Clarissa an SMS message saying

> Found graph, 10 vert's, 24 edg's, 8 triang's, 16 quad's, no other cyc's.

At first, Clarissa is amazed, but then she concludes that this graph must have a subgraph with 10 edges that has no cycles at all. Now she is amused. Why could she come to the conclusion, and why is she amused?

5 Expected Height of Random Search Trees

A *random search tree* for a set S of n keys (say, natural numbers, pairwise distinct) is recursively built as follows: If S is empty, then the tree is empty (i.e. it has no vertex at all). Otherwise, we choose a random key w from S (each key in S with probability $\frac{1}{n}$), we create a root vertex which stores w, and two pointers to a left child (which is a random search tree for $S_{<w} := \{k \in S \mid k < w\}$), and a right child (which is a random search tree for $S_{>w} := \{k \in S \mid k > w\}$); if $S_{<w}$ or $S_{>w}$ is empty, then the respective child does not exist. The following figure displays all search trees for three keys, each with its probability to occur as a random search tree.



A number of expected quantities are quite easy to derive for such a random search tree (if you know how). For example the expected depth of the smallest key in the tree is $H_n - 1$ (this closely related to the number of left-to-right minima in a random permutation); in general, the expected depth of the k-smallest key (we call this the key of *rank* k) is $H_{n-k+1} + H_k - 2$. The expected sum of the depth of all n keys is $2(n+1)H_n - 4n$.

Here we are interested in the expected maximum depth of such a tree, i.e. the height of a random search tree for n keys. The following describes a proof of an upper bound for this quantity – surprisingly enough (after having seen the proof), this bound is already tight as for the constant in the leading term.

So, for $1 \le i \le n$, let $Y_n^{(i)}$ be the random variable for the depth of the key of rank i in a random search tree for n keys. Then $X_n := \max\{Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(n)}\}$ is the random variable for the height of the tree with n keys. Analyzing the maximum of random variables is usually quite difficult (even more so than higher moments). But we will find a way around.

We can write

$$\mathrm{E}(X_n) \le \lg\, \mathrm{E}\Big(2^{X_n}\Big) = \lg\, \mathrm{E}\Big(2^{\max_{1 \le i \le n} Y_n^{(i)}}\Big) < \lg\, \mathrm{E}\left(\sum_{1 \le i \le n,\ i \text{ is leaf}} 2^{Y_n^{(i)}}\right). \qquad (6)$$

The first inequality uses Jensen's inequality which states that $f(\mathrm{E}(X)) \le \mathrm{E}(f(X))$ for any convex function f (provided the expectations exist). We can now estimate a random variable involving '$\max$' by the random variable $Z_n := \sum_{1\le i\le n,\ i\text{ is leaf}} 2^{Y_n^{(i)}}$ involving a sum. The conditional part 'i is leaf' causes no problems as we shall see; in fact, we put it there to make life easier.

Obviously, $\mathrm{E}(Z_0) = 0$ and $\mathrm{E}(Z_1) = 1$. There are two trees on two vertices, each one appears with probability $\frac{1}{2}$, which gives $\mathrm{E}(Z_2) = 2$; by checking the figure above for the case of three keys, we obtain $\mathrm{E}(Z_3) = 4$. Enough of fiddling around with small values. For $n \ge 2$, we get

$$\mathrm{E}(Z_n) = \frac{1}{n}\sum_{1\le i\le n} \mathrm{E}(Z_n \mid \text{root has rank } i)\ ,$$

where $\mathrm{E}(Z_n \mid \text{root has rank } i) = 2(\mathrm{E}(Z_{i-1}) + \mathrm{E}(Z_{n-i}))$. Setting $z_n := \mathrm{E}(Z_n)$, we have $z_0 = 0$, $z_1 = 1$ and, for $n \ge 2$,

$$z_n = \frac{4}{n}\sum_{0\le j\le n-1} z_j\ . \tag{7}$$

Consequently, for $n \ge 3$, $nz_n - (n-1)z_{n-1} = 4z_{n-1}$, and so $nz_n = (n+3)z_{n-1}$ or

$$\frac{z_n}{(n+3)(n+2)(n+1)} = \frac{z_{n-1}}{(n+2)(n+1)n} = \cdots = \frac{z_2}{5\cdot 4\cdot 3} = \frac{1}{30}\ .$$

That is, $z_n = \frac{(n+3)(n+2)(n+1)}{30}$ for $n \ge 2$.

If we plug that bound into (6), then we get an upper bound of $3\lg n + O(1) = 4.32808\ldots \ln n + O(1)$ for the expected height of a random search tree. Does this give already the right constant? Well, contrary to what we announced before, this is not the case – *yet*! But we still have the base 2 to play with.

We set $Z_n := \sum_{1\le i\le n,\ i\text{ is leaf}} C^{Y_n^{(i)}}$, with C some real number greater than 1. Similar to (6), we have $\mathrm{E}(X_n) \le \log_C \mathrm{E}(Z_n)$. The recursion for $z_n := \mathrm{E}(Z_n)$ is the same as the one given in (7), except that 4 gets replaced by $2C$, which eventually leads to $nz_n = (n + 2C - 1)z_{n-1}$ for $n \ge 2$. This yields

$$z_n = (1 + \frac{2C-1}{n})z_{n-1} = (1 + \frac{2C-1}{n})(1 + \frac{2C-1}{n-1})\cdots(1 + \frac{2C-1}{3})z_2\ .$$

Since $z_2 = C \le (1 + \frac{2C-1}{2})$, and $1 + x \le \mathrm{e}^x$ for any real number x, this implies

$$z_n \le \mathrm{e}^{(2C-1)\sum_{i=2}^{n}(1/i)} = \mathrm{e}^{(2C-1)(H_n-1)} < \mathrm{e}^{(2C-1)\ln n} = n^{2C-1}\ . \tag{8}$$

Invoking (6) (in its adopted version with 'C' instead of '2'), this gives a bound of

$$\mathrm{E}(X_n) < \frac{2C-1}{\ln C}\ln n\ .$$

This bound attains its extremal values if $2\ln C - 2 + 1/C = 0$, or, equivalently, if $(\frac{\mathrm{e}}{C})^{2C} = \mathrm{e}$. Note that for these values of C, we have $\frac{2C-1}{\ln C} = 2C$. Hence, we have shown

Theorem 5.1 *The expected height of a random search tree for n keys is bounded by $c\ln n$, where $c = 4.311070\ldots$ is the unique value greater than 2 which satisfies $(\frac{2\mathrm{e}}{c})^c = \mathrm{e}$.* □

The constant in the leading term is already tight, as it was shown by Devroye; the proof of this fact is considerably more involved than the upper bound proof we have just seen. The upper bound has been shown before by Robson.

Let us conclude by pointing out that knowing a good estimate for $\mathrm{E}\left(C^{X_n}\right)$ immediately gives a good tail estimate for X_n via Markov's Inequality, namely

$$\Pr(X_n > \tau\ln n) = \Pr\left(C^{X_n} > C^{\tau\ln n}\right) = \Pr\left(C^{X_n} > \frac{C^{\tau\ln n}}{\mathrm{E}(C^{X_n})}\mathrm{E}\left(C^{X_n}\right)\right) \le$$

$$\frac{\mathrm{E}\left(C^{X_n}\right)}{C^{\tau\ln n}} < n^{2C-1-\tau\ln C}\ .$$

$2C - 1 - \tau\ln C$ is minimized for $C = \tau/2$.

Theorem 5.2

$$\Pr(X_n > \tau\ln n) < n^{\tau(1-\ln(\tau/2))-1}\ ;$$

in particular $\Pr(X_n > 2\mathrm{e}\ln n) < \frac{1}{n}$.

If we set $\tau = c$, where c is the constant in Theorem 5.1, then we obtain $\Pr(X_n > c\ln n) < 1$, which we might have guessed before. However, you may want to reconsider the estimate obtained in (8) to get a nontrivial (though still constant) bound for this probability.

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).

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, $k, n \in \mathbb{N}_0$, 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

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

Note the similarity to the recurrence for binomial coefficients,

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

and how $\sum_{i=0}^{n} \genfrac{[}{]}{0pt}{}{n}{i} = n!$ 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

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

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

Exercise 6.1 *Prove the recurrence*

$$\genfrac{[}{]}{0pt}{}{n}{k} = (n-1)\genfrac{[}{]}{0pt}{}{n-1}{k} + \genfrac{[}{]}{0pt}{}{n-1}{k-1} \ .$$

Do it once reading $\genfrac{[}{]}{0pt}{}{n}{k}$ *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*

$$\genfrac{[}{]}{0pt}{}{n}{1}, \genfrac{[}{]}{0pt}{}{n}{2}, \genfrac{[}{]}{0pt}{}{n}{n-1}, \ \textit{and} \ \genfrac{[}{]}{0pt}{}{n}{n} \ , n \in \mathbf{N}.$$

Exercise 6.3 Sums of Stirling Cycle Numbers

What can you say about

$$\sum_{i=0}^{n} i \genfrac{[}{]}{0pt}{}{n}{i} \ , \quad \sum_{i=0}^{n} i^2 \genfrac{[}{]}{0pt}{}{n}{i} \ , \ \textit{and} \ \sum_{i=0}^{n} 2^i \genfrac{[}{]}{0pt}{}{n}{i} \ \ ?$$

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*

$$\sum_{i=0}^{n} (-1)^i \genfrac{[}{]}{0pt}{}{n}{i} = 0 \ ?$$

Exercise 6.8 Estimating the Stirling Cycle Numbers

For $i, n \in \mathbf{N}_0$, $i \le n$, *show that*

$$\genfrac{[}{]}{0pt}{}{n+1}{i+1} \le \frac{n!}{i!}(H_n)^i \ .$$

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

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

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

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

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

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 $\left(1-\frac{1}{n}\right)^n < \frac{1}{\mathsf{e}}$.

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

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

For $k=n$, we expect to see roughly $n(1-\frac{1}{\mathsf{e}}) = 0.63\ldots n$ 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

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

[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, $\mathrm{E}(Y_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$. We sum up these expectations for $i = 1, 2, \ldots, n$ to obtain

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

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 $\mathrm{E}(X^2)$ 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 $\mathrm{E}(W_i)$, $i \in \{1..n\}$ and $\mathrm{E}\Big(\max_{i\in\{1..n\}} W_i\Big)$?*

Exercise 7.4 *Draw a square (about $10\mathrm{cm} \times 10\mathrm{cm}$) 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.

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

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. $\square$

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 $G_{\text{red}} = (V, E_{\text{red}})$, the red subgraph, and $G_{\text{blue}} = (V, E_{\text{blue}})$, 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,

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

[11] Almost verbatim from [6].

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

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. $\boxminus$

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

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

◰

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*

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

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

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

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

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

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
    $a \leftarrow_{\mathrm{random}} G$;
    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($A$)
$a \leftarrow_{\mathrm{random}} A$;
**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, $0 \le k \le n$, recursively defined by $f(k,k) = 0$ for $k \in \mathbf{N}_0$ and

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

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