Contents

0.1 Probability 1
  0.1.1 Some Experiments 1
  0.1.2 Basic Concepts 3
  0.1.3 Counting 5
  0.1.4 Random Variables 10
  0.1.5 Chernoff Bounds 14
  0.1.6 Conditional Probability and Independence 14
  0.1.7 Bayes's Rule 14
  0.1.8 Continuous Probability Spaces 14

0.1 Probability

Probability is about chance experiments, like flipping a fair coin or rolling a dice or playing poker or running an algorithm which flips coins during its execution. Probability theory provides a language to talk and to argue about chance experiments. Probability is ubiquitous in the analysis of algorithms other parts of computer science and therefore we need a working knowledge of basic probability theory.

0.1.1 Some Experiments

We give some examples of chance experiments. *The computer is a great tool in studying discrete probability spaces and random variables.* We urge our students to perform experiments. The authors almost always run experiments when they have to analyze a probabilistic quantity.

Throwing Balls into Bins: The balls into bins experiment lies at the heart of the analysis of hashing and of many storage allocation problems in parallel computing and hierarchical memory.

Consider the following experiment which is usually called "throwing n balls randomly into m bins". We generate n pseudo-random number in the range $[0 .. m-1]$ and count the occurrences of each integer. We report the maximal number of occurrences of any integer. The experiment is repeated k times.

⟨*balls_into_bins.c*⟩≡

```
#include <LEDA/array.h>
#include <LEDA/random_source.h>
int main(){
```

```
int n = read_int("number of balls = ");
int m = read_int("number of bins  = ");
int K = read_int("number of repetitions = ");
double Xmax_ave = 0, X0_ave = 0;
for (int k = 0; k < K; k++)
{ array<int> X(m);
  X.init(0);  // initializes all entries to zero
  for (int i = 0; i < n; i++)
  { int j = rand_int(0,m-1);  // a random integer in [0 .. m - 1]
    X[j]++;
  }
  int Xmax = 0;
  for (int i = 0; i < m; i++) Xmax = leda_max(Xmax,X[i]);
  cout << "Xmax = " << Xmax << "\n";
  Xmax_ave += Xmax; X0_ave += X[0];
}
newline;
cout << "The average value of Xmax is equal to " << Xmax_ave/K << ".\n";
cout << "The average value of X[0] is equal to " << X0_ave/K << ".\n";
}
```

A sample run produced

```
number of balls = 1000
number of bins  = 1000
number of repetitions = 5
Xmax = 5
Xmax = 5
Xmax = 6
Xmax = 6
Xmax = 5
The average value of Xmax is equal to 5.4.
The average value of X[0] is equal to 1.3.
```

Table 1 gives more data.

Left to Right Maxima: The following problem arises in the analysis of sorting algorithms. Consider an arbitrary sequence $(x_1, x_2, \ldots, x_n)$ of numbers. We call the i-th number in the sequence a left-to-right maximum if x_i is at least as large as all preceding numbers. What is the number of left to right maxima in a random permutation of the integers 0 to $n-1$. The following program gives a feeling.

⟨*left_to_right_maxima.c*⟩≡

```
#include <LEDA/array.h>
int main(){
int n = read_int("n = ");
int K = read_int("number of repetitions = ");
array<int> A(n);
double number_of_max_ave = 0;
for (int k = 0; k < K; k++)
{ for (int i = 0; i < n; i++) A[i] = i;
```

	1000		2000		4000		8000	
1000	5.6	1.03	7.83	1.88	11.83	4.15	18.56	7.76
2000	4.36	0.46	5.84	1	8.38	1.88	12.56	3.84
4000	3.47	0.25	4.66	0.43	6.21	1.13	8.87	1.99
8000	2.95	0.14	3.78	0.29	4.75	0.55	6.58	0.99
16000	2.55	0.11	3.08	0.1	4.07	0.23	5.21	0.37
32000	2.18	0.04	2.77	0.07	3.19	0.23	4.21	0.25
64000	2.07	0.03	2.24	0.04	2.97	0.06	3.41	0.09
128000	1.95	0	2.05	0.02	2.55	0.05	3.02	0.04

Table 1: The rows correspond to different values of m and columns correspond to different values of n. For each n the first column gives the average value of *Xmax* and the second column shows the average value of $X[0]$. All values are averages of 100 runs. *Try to find the law behind these numbers.*

$n/100 =$	1	2	4	8	16	32	64	128	256	512	1024	2048
maxima	5.07	6.03	6.81	7.94	7.91	9.2	9.25	10.7	10.69	12.03	12.45	12.74

Table 2: The number of left to right maxima in a random permutation of n elements. All numbers are the average of 100 runs. *Try to find the law behind these numbers.*

```
    A.permute();  // permutes A randomly
    int max = -1, number_of_max = 0;
    for (int i = 0; i < n; i++)
      if ( A[i] > max ) { max = A[i]; number_of_max++; }
    number_of_max_ave += number_of_max;
  }
  cout <<"average number of left right maxima = " << number_of_max_ave/K << "\n\n";
  }
```

Run this program for different values of n and form a conjecture for the average number of left to right maxima. Table 2 gives some more data.

Observe that the number of left-to-right maxima actually tells us something about the running time of the program above. It is the number of times that we are going to execute the body of the conditional `if ( A[i] > max )`.

0.1.2 Basic Concepts

Probability theory explains chance experiments. We start with the basic concepts sample space, event, and probability distribution. A *sample space* S is simply a set. The elements of a sample space are called *elementary events* and some of the subsets of S are called *events*. Elementary events are viewed as the possible outcomes of an experiment.

For the balls-into-bins experiment we take $S = [0 \mathrel{..} m-1]^n$, i.e., all sequences of length n formed from integers between 0 and $m-1$, and for the left-to-right-maxima experiment we take S to be the set of permutions of the integers 0 to $n-1$.

A probability space is called *discrete* if S is a finite or countably infinite set and it is called *continuous* otherwise. We will almost exclusively deal with discrete sample spaces. Certain subsets of S are called *events*. In the case of a discrete probability space all subsets of S are events and hence we can postpone the distinction between event and subset of S until we deal with continuous probability spaces in Section 0.1.8. Two events A and B are called *mutually exclusive* if they they are disjoint, i.e., if $A \cap B = \emptyset$ and two events are called *complementary* if $A \cap B = \emptyset$ and $A \cup B = S$. If A and B are complementary, we also write $A = \overline{B}$. The empty set $\emptyset$ is called the *null event*.

A *probability distribution* prob( ) on a discrete sample space S assigns a non-negative real number $\text{prob}(s)$ or p_s to every elementary event $s \in S$ such that $\sum_{s\in S} p_s = 1$. We extend prob( ) to arbitrary events $E \subseteq S$ via

$$\text{prob}(E) = \sum_{s\in E} p_s$$

and call $\text{prob}(E)$ the probability of event E. Let us make sure that we understand the definition by proving a few simple facts.

- The probability of the null event is zero. Indeed, $\sum_{s\in\emptyset} p_s = 0$.
- If A is a subset of B then $\text{prob}(A) \le \text{prob}(B)$. Indeed, $\text{prob}(A) = \sum_{s\in A} p_s \le \sum_{s\in B} p_s = \text{prob}(B)$.
- If A and B are arbitrary events, then $\text{prob}(A \cup B) + \text{prob}(A \cap B) = \text{prob}(A) + \text{prob}(B)$. Indeed, elements in $(A \cup B) \setminus (A \cap B)$ contribute once to both sides and elements in $A \cap B$ contribute twice to each side.
- If A and B are mutually exclusive, then $\text{prob}(A) + \text{prob}(B) = \text{prob}(A \cup B)$. This follows from the preceding item and the the fact that $\text{prob}(()A \cap B) = \text{prob}(()\emptyset) = 0$ for mutually exclusive events.
- $\text{prob}(\overline{A}) = 1 - \text{prob}(A)$ for all events A. This follows from the preceding item and the fact that $\text{prob}(()A \cup B) = \text{prob}(()S) = 1$ for complementary events..

A particularly important distribution on finite sample spaces S is the *uniform distribution* which assigns probability $1/|S|$ to each elementary event. The probability of an arbitrary event E is then simply the fraction $|E|/|S|$.

We will use uniform probability spaces (= a finite probability space with the uniform distribution defined on it) extensively. We will use the space $S = \{H, T\}^n$ with the uniform distribution to model the experiment of throwing n fair coins[1]. The size of S is 2^n and hence the probability of each elementary event is 2^{-n}. The probability of an arbitrary event is simply $|E| \cdot 2^{-n}$. Let us compute some probabilities:

- What is the probability of no heads in a sequence of n coin tosses? This is the probability of the event $\{TT \ldots T\}$. The cardinality of this event is equal to one and hence its probability is 2^{-n}.
- What is the probability of exactly k heads in a sequence of n coin tosses. This the probability of the event

$$E = \{\text{all sequences of length } n \text{ over } \{H, T\} \text{ containing exactly } k\ T\text{'s}\}\ .$$

[1] A coin is called *fair* if the probability of coming up heads is the same as the probability of coming up tails, i.e., both probabilities are equal to $1/2$.

The cardinality of E is $\binom{n}{k}$ and hence the probability of E is $\binom{n}{k} \cdot 2^{-n}$. For $k = 0$ this is the result of the previous item.

- What is the probability that exactly k balls end up in the first bin after throwing n balls into m bins?

Computing probabilities in uniform sample spaces frequently amounts to counting combinatorial objects with a certain property.

0.1.3 Counting

We review some basic facts about permutations, strings, and binomial coefficients.

Permutations: A *permutation* of a finite set S is a sequence in which every element of S appears exactly one. There are $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ permutations of an n element set. Indeed, there are n ways to choose the first element, $n-1$ ways to choose the second element, $n-2$ ways to choose the third element, ..., two ways to choose the next to last element, and only one way to choose the last element. $n!$ is read as "n factorial". We define $O!$ as 1. We will see the usefulness of this definition soon.

We frequently want to estimate $n!$. The bounds $n! \le n^n$ (every factor in the product expansion of $n!$ is at most n) and $n! \ge (n/2)^{n/2}$ (at least half of the factors are at least $n/2$) are trivial. *Stirling's approximation* states that the average factor in the product expansion of $n!$ is about n/e.

Theorem 1 (Stirling's Approximation)

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

Proof: We prove slightly weaker bounds. For a proof of the stated bounds we refer the reader to [Knu73, page 111].

The proof has a simple structure. We write $n!$ as $e^{\ln n!}$ and then estimate $\ln n!$ by an integral. We have

$$\ln n! = \ln \prod_{i=1}^{n} i = \ln \prod_{i=2}^{n} i = \sum_{i=2}^{n} \ln i .$$

Next observe that

$$\int_{i-1}^{i} \ln x \, \mathrm{d}x \le \ln i \le \int_{i}^{i+1} \ln x \, \mathrm{d}x$$

since ln is an increasing function. Thus

$$\int_{1}^{n} \ln x \, \mathrm{d}x \le \ln n! = \sum_{i=2}^{n} \ln i \le \int_{2}^{n+1} \ln x \, \mathrm{d}x .$$

Since $\int \ln x \, \mathrm{d}x = x(\ln x - 1) = x \ln(x/e)$ we conclude further

$$x \ln(x/e)|_1^n \le \ln n! \le x \ln(x/e)|_2^{n+1}$$

or

$$n \ln(n/e) \leq \ln n! \leq (n+1)\ln((n+1)/e) - 2\ln(2/e)$$

or

$$\ln((n/e)^n) \leq \ln n! \leq \ln((e/2)^2 \cdot ((n+1)/e)^{n+1})$$

or

$$(n/e)^n \leq n! \leq 2 \cdot ((n+1)/e)^{n+1} \; .$$

∎

Permutations of an n-element set can also be viewed as bijections between the integers from 1 to n and the set S. We simply identify the i-th element in the sequence with the image of i.

Ordered sequences without repetitions generalize permutations. The number of repetition free sequences of length k that can be formed from an n element set is equal to

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!} \; .$$

Indeed, there are n ways to choose the first element, $n-1$ ways to choose the second element, ..., and $n-k+1$ choices for the last element.

Exercise:

1. TODO

Binomial Coefficients and their Properties *Binomial coefficients* are ubiquitous in counting problems. The coefficient $\binom{n}{k}$ (read "n choose k") is defined as

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

By virtue of the definition $0! = 1$, $\binom{n}{0} = \binom{n}{n} = 1$, and binomial coefficients apply also for $k = 0$ and $k = n$. For $k > n$, we define $\binom{n}{k} = 0$. Binomial coefficients count the number of k-element subsets of an n-element set.

Lemma 1 *The number of subsets of size k of an n-element set is equal to $\binom{n}{k}$.*

Proof: We give two proofs. Let S be any set of size n. For the first proof we define a mapping from permutations of S to subsets of S of size k. We map a permutation to the set formed by its first k elements.

For example, for $n = 4$ and $k = 2$ and $S = \{1, 2, 3, 4\}$ the permutation $(2, 3, 1, 4)$ is mapped to the set $\{2, 3\}$ and so are the permutations $(3, 2, 1, 4)$, $(2, 3, 4, 1)$ and $(3, 2, 4, 1)$. Thus $4 = 2! \cdot 2!$ permutations map to the same set.

In general, $k! \cdot (n-k)!$ permutations map to the same set, since we may arbitrarily reorder the first k and the last $n-k$ elements without changing the induced set. Thus every set is produced by $k! \cdot (n-k)!$ permutations and hence the number of sets is $\frac{n!}{k!\cdot(n-k)!}$.

For the second proof we use induction on k. There is exactly one 0-element subset of S, namely the empty set. Also $1 = \binom{n}{0}$.

Assume now $k > 0$. Induction hypothesis tells us that there are $\binom{n}{k-1}$ subsets of S of size $k-1$. We can turn a subset of size $k-1$ into a k element subset by adding any one of the remaining $n-(k-1)$ elements. In this way every k-element subset is obtained k times, since any of its elements could be the element added last. We conclude that the number of k-element subsets is equal to

$$\frac{n-k+1}{k}\binom{n}{k-1} = \frac{(n-k-1)n!}{k(k-1)!(n-k+1)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k} .$$

∎

There are numerous identities involving binomial coefficients. We give some examples. All of them follow by short calculations from the definition of binomial coefficients:

$$\begin{aligned}
\binom{n}{k} &= \binom{n}{n-k} , \\
\binom{n}{k} &= \frac{n}{k}\binom{n-1}{k-1} \text{ for } 0 < k \le n , \\
\binom{n}{k} &= \frac{n}{n-k}\binom{n-1}{k} \text{ for } 0 \le k < n , \\
\binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1} \text{ for } 0 < k \le n ,
\end{aligned}$$

The identities can also be obtained by combinatorial reasoning. Combinatorial reasoning gives more insight and is more fun than brute-force calculation. Combinatorial reasoning amounts to counting the size of a set in two different ways.

For the first identity, we observe that the mapping $A \mapsto S \setminus A$ maps any k-element subset A to a $(n-k)$-elements subset.

For the second identity, we count the set of pairs (A, x) with A being a subset of S of size k and x being an element of A. We can generate such pairs by first selecting A ($\binom{n}{k}$ possibilities) and then choosing x in A (k possibilities) or by first selecting x (n possibilities) and then selecting the remaining elements of A from the remaining elements of S ($\binom{n-1}{k-1}$ possibilities).

For the third identity, we count the set of pairs (A, x) with A being a subset of S of size k and x being an element of $S \setminus A$. We can generate such pairs by first selecting A ($\binom{n}{k}$ possibilities) and then choosing x in $S \setminus A$ ($n-k$ possibilities) or by first selecting x (n possibilities) and then selecting A from the remaining elements of S ($\binom{n-1}{k}$ possibilities).

For the last identity, we consider a set of n elements of which one is red and the other are green. A subset of size k either consists only of green elements or it contains the red element. There are $\binom{n-1}{k}$ sets of the former kind and $\binom{n-1}{k-1}$ sets of the latter kind.

There is a second very useful combinatorial interpretation of binomial coefficients: $\binom{n}{k}$ counts the number of strings over the binary alphabet $\{0, 1\}$ of length n containing exactly k zeroes. Indeed, if $x_1x_2\dots x_n$ is any such string, $A = \{i\ ;x_i = 0\}$ is a subset of $[0\,..\,n-1]$ of size k and this mapping from strings to sets is bijective. The alternative combinatorial interpretation gives us a proof of the *binomial theorem*

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

which expresses the n-th fold power of two variables x and y as a sum involving binomial coefficients. The equality can be verified as follows. Expanding $(x+y)^n$ yields all strings of length n over the alphabet $\{x, y\}$. The number of strings containing i x's and $n-i$ y's is equal to $\binom{n}{i}$.

With $x = y = 1$ the binomial theorem yields the remarkable identity

$$2^n = \sum_i \binom{n}{i}\ ,$$

which can also be verified combinatorially by counting the number of strings over the alphabet $\{0, 1\}$ in two different ways. There are 2^n such strings. The right hand side counts the number of strings according to the number of zeroes contained in them.

We have now seen several impressive identities which have simple combinatorial explanations. Here is another one.

$$\binom{2n}{k} = \sum_{0\le i\le k} \binom{n}{i}\binom{n}{k-i}\ .$$

Both sides count the number of strings of length $2n$ over $\{0, 1\}$ containing k zeroes. Any such string has i zeroes in its first half and $k-i$ zeroes in its second half for some i and this is the way the strings are counted by the expression on the right. Impressed?

We also need estimates for binomial coefficients. TODO give table

Lemma 2 *For $1 \le k \le n$ we have*

$$(n/k)^k \le \binom{n}{k} \le (en/k)^k\ .$$

Proof: For the lower bound we observe that

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} \ge (n/k)^n\ ,$$

since $(n-i)/(k-i) \ge n/k$ for $0 \le i < k$.

For the upper bound we use the inequality $k! \ge (k/e)^k$ implied by Stirling's approximation and obtain

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} \le \frac{n^k}{k!} \le (en/k)^n\ .$$

∎

Words and Functions: A *word* or *string* of length n over an alphabet X is a sequence of n elements of X, i.e, a sequence $x_1 \dots x_n$ with $x_i \in X$. Clearly, if $|X| = m$, there are m^n words of length n. We considered the special case $m = 2$ already in the previous paragraph.

We can view words also as functions from $[n]$ to $[m]$. More precisely, a word $x = x_1 \dots x_n$ defines a function f with $f(i) = x_i$ and a function f from $[n]$ to $[m]$ defines a word by setting $x_i = f(i)$. Thus there are m^n functions from $[n]$ to $[m]$.

The number of functions without a fix-point, i.e., $f(i) \neq i$ for all i is $(m-1)^n$. Indeed, we have only $m-1$ possibilities for each function value.

The number of functions with no duplicate values is $m(m-1)\cdots(m-n+1)$ as we know already from Section **??**.

Exercises:

1. Consider the following directed graph. The nodes are the integer lattice points (i, j) with $i \in \mathbb{N}_0$ and $j \in \mathbb{N}_0$. The edges are either horizontal and to the right or diagonal and to the right, i.e., from the point (i, j) we have the horizontal edge to $(i+1, j)$ and the diagonal edge to $(i+1, j+1)$.
   - Argue that the number of paths from $(0, 0)$ to (n, k) is equal to $\binom{n}{k}$.
   - Let m be an arbitrary integer between 0 and n and let i be an arbitrary integer between 0 and k. Argue that there are exactly $\binom{m}{i}\binom{n-m}{k-i}$ paths from $(0, 0)$ to (n, k) which pass through the point (m, i).
   - Fix m. Any path from $(0, 0)$ to (n, k) goes through exactly one point whose first coordinate is equal to m. Use this observation and the previous exercise to verify the identity
   $$\binom{n}{k} = \sum_i \binom{m}{i}\binom{n-m}{k-i}.$$
   - Consider the following variant of the second item. We now count paths which pass through the points $(m-1, i-1)$ and (m, i).
   - Fix i. For any path from $(0, 0)$ to (n, k) there is exactly one value of m such that the path passes through the points $(m-1, i-1)$ and (m, i). Use this observation and the previous exercise to verify the identity
   $$\binom{n}{k} = \sum_m \binom{m-1}{i-1}\binom{n-m}{k-i}.$$
2. Prove the identity $\sum_{0 \le i \le n} i = \binom{n+1}{2}$ by induction on n.
3. Prove the identity
   $$\binom{n}{k} = \sum_{k-1 \le m < n} \binom{m}{k-1},$$
   by arguing that both sides count the number of binary strings of length n containing exactly k zeroes. Group strings according to the position of the last zero. Specialize to the case $k = 2$. What is the relation to the first exercise?

0.1.4 Random Variables

We now come to an extremely useful concept, the concept of a *random variable*. It is simply a function from a probability space into the real numbers. We (as most other people do) use capital letters near the end of the alphabet to denote random variables. X and Y will be our canonical notation for random variables.

Here are some first examples.

- $S = \{H, T\}^n$ and X is the number of heads, i.e., for $s = s_1 s_2 \ldots s_n \in \{H, T\}^n$ we set $X(s)$ equal to the number of i with $s_i = H$. Instead of "$S = \{H, T\}^n$ and X is the number of heads" we will even say more succinctly "X is the number of heads in n tosses of a fair coin".

- $S = [1 \mathinner{..} m]^n$ and for $s = s_1 s_2 \ldots s_n \in [1 \mathinner{..} m]^n$ and any $j \in [1 \mathinner{..} m]$, we set $X_j(s)$ equal to the number of i with $s_i = j$. The probability space can either be interpreted as rolling n dices with m sides each or throwing n balls into m bins; X_j is the number of dices coming up with the number j and the number of balls in bin j, respectively. We will also talk about the occupancy of bin j. The random variable $X = \max(X_1, \ldots, X_m)$ is also of interest. It measures the maximal occupancy of any bin.

Random variables can be used to define events. The event that the value of X is equal to x, in short "the event $X = x$", is the set $\{s \in S\,;\, X(s) = x\}$. Its probability is

$$\mathrm{prob}(X = x) = \sum_{\{s \in S\ ;\ X(s) = x\}} p_s\ .$$

The *expectation* of a random variable is its "average value".

Definition 1 *The expectation* $\mathrm{E}[X]$ *of a random variable* X *is defined as*

$$\mathrm{E}[X] = \sum_{s \in S} p_s \cdot X(s)\ .$$

There is a alternative way of computing the expectation. We may alternatively sum over the values in the range of the random variable.

Lemma 3 $\mathrm{E}[X] = \sum_{x \in \mathbb{R}} \mathrm{prob}(X = x) \cdot x$.

Proof: We prove the equality by rearranging sums. Let R be the range of X, i.e., $R = \{X(s)\,;\, s \in S\}$. R is a countable subset of $\mathbb{R}$ and hence there is no danger in summing over the elements in R. We have

$$\begin{aligned}
\sum_{s \in S} p_s \cdot X(s) &= \sum_{x \in R} \sum_{s \in S \text{ and } X(s) = x} p_s \cdot x \\
&= \sum_{x \in R} x \cdot \sum_{s \in S \text{ and } X(s) = x} p_s \\
&= \sum_{x \in R} x \cdot \mathrm{prob}(X = x) \\
&= \sum_{x \in \mathbb{R}} x \cdot \mathrm{prob}(X = x),
\end{aligned}$$

where the last equality follows from the fact that $\mathrm{prob}(X = x) = 0$ for $x \notin R$. ∎

For integral valued random variables there is yet another way to compute expectations.

Lemma 4 *The X be a random variables whose values are non-negative integers. Then*

$$\mathrm{E}[X] = \sum_{x>0} \mathrm{prob}(X \geq x) = \sum_{x \geq 0} \mathrm{prob}(X > x) .$$

Proof: We have $\mathrm{prob}(X = x) = \mathrm{prob}(X \geq x + 1) - \mathrm{prob}(X \geq x)$ and hence

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{x \geq 0} \mathrm{prob}(X = x) \cdot x \\
&= \sum_{x \geq 0} (\mathrm{prob}(X \geq x + 1) - \mathrm{prob}(X \geq x)) \cdot x \\
&= \sum_{x \geq 0} \mathrm{prob}(X \geq x + 1) \cdot (x - (x - 1)) \\
&= \sum_{x \geq 0} \mathrm{prob}(X \geq x + 1) \\
&= \sum_{x > 0} \mathrm{prob}(X \geq x)
\end{aligned}$$

∎

Let us compute some expectations. What is the expected value in a single toss of a fair dice? Our probability space is $\{1, 2, \dots, 6\}$ equipped with the uniform distribution and we consider the random variable X defined by $X(s) = s$ for $1 \leq s \leq 6$. We obtain for the expectation

$$\mathrm{E}[X] = 1/6 \cdot 1 + 1/6 \cdot 2 + \ldots + 1/6 \cdot 6 = 7/2 .$$

What is the expected sum in 2 tosses of a fair dice? Our probability space is $\{1, 2, \dots, 6\} \times \{1, 2, \dots, 6\}$. There is a single way to obtain the value 2, there are two ways to get the number 3, there are three ways to get the number 4, ..., there are five ways to get the number 6, there are six ways to get the number seven, there are five ways to get the number 8, ..., there is one way to get the number 12. Thus

$$\begin{aligned}
\mathrm{E}[\quad] = &\ (1 \cdot 2 + 2 \cdot 3 + \ldots + 5 \cdot 6 + 6 \cdot 7 + 5 \cdot 8 + \ldots + 1 \cdot 12)/36 \\
= &\ ((1 + 2 + 3 + 4 + 5) \cdot 14 + 42)/36 \\
= &\ (15 \cdot 14 + 42)/36 = (35 + 7)/6 = 7 .
\end{aligned}$$

This was hard work. We will soon learn a much better way of estimating this expectation. Try to compute the expected sum of four fair dices.

The following theorem is extremely useful (and trivial to prove).

Theorem 2 (Linearity of Expectations) *Let X and Y be two random variables on the same sample space S. Then*

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y] .$$

Proof: We have

$$\begin{aligned} \mathrm{E}[X+Y] &= \sum_{s\in S} p_s \cdot (X(s)+Y(s)) \\ &= \sum_{s\in S} p_s \cdot X(s) + \sum_{s\in S} p_s \cdot Y(s) \\ &= \mathrm{E}[X] + \mathrm{E}[Y] \,. \end{aligned}$$

∎

Warning: The theorem above does not hold when summation is replaced by multiplication.

Let us go back to our problem of computed the expected sum of the tosses of two fair dices. Let X_1 be the value of the first dice and let X_2 be the value of the second dice. Then $\mathrm{E}[X_1] = \mathrm{E}[X_2] = 7/2$ and hence $\mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 7$. Isn't it amazing how simple things get, if one has the right tools.

We next turn to our ball-into-bins experiment. Let X_j be the number of balls thrown into the j-th bin. What is the expectation of X_j? We give three ways of computing it: two elegant ways and one inelegant way. The inelegant way is the most natural one. Let $p_{j,l}$ be the probability that l balls are thrown in bin j. Then $X_j = \sum_l p_{j,l} \cdot l$. For any individual ball the probability of ending up in bin j is $1/m$ and the probability of not ending up in bin j is $1 - 1/m$. Also there are $\binom{n}{l}$ ways of selecting the l balls that end up in bin j. Thus

$$p_{j,l} = \binom{n}{l}(1/m)^l(1-1/m)^{n-l}$$

and hence

$$\begin{aligned} \mathrm{E}[X_j] &= \sum_{l\ge 0} \binom{n}{l}(1/m)^l(1-1/m)^{n-l} \cdot l \\ &= \frac{n}{m} \sum_{l>0} \binom{n-1}{l-1}(1/m)^{l-1}(1-1/m)^{n-l} \\ &= n/m \,, \end{aligned}$$

where the first equality uses the identity $l\binom{n}{l} = n\binom{n-1}{l-1}$ for $l > 0$ and the second equality follows from the binomial theorem.

We come to our second proof. We have $X_0 + X_1 + \ldots + X_{m-1} = n$ since n balls are thrown and any balls ends up in some bin. Thus by linearity of expectations $n = \sum_{0\le j<m} \mathrm{E}[X_j]$. Symmetry tells us the $\mathrm{E}[X_j]$ is the same for all j. Thus $\mathrm{E}[X_j] = n/m$ for all j. We urge the reader to think about the phrase "symmetry tells us" for at 15 minutes. Try to expand the phrase to 10 lines.

And here comes our third proof. Define the indicator variables[2] $Y_{i,j}$ by $Y_{i,j} = 1$ if the i-th ball ends up in bin j and 0, otherwise. The $X_j = \sum_{0\le i<n} Y_{i,j}$ and hence by

[2] A random variable which takes only values zero and one is called an *indicator variable*. For an indicator variable X we have

$$\mathrm{E}[X] = \mathrm{prob}(X=0)\cdot 0 + \mathrm{prob}(X=1)\cdot 1 = \mathrm{prob}(X=1) \,,$$

a very useful identity.

linearity of expectations

$$\mathrm{E}[X_j] = \sum_{0 \le i < n} \mathrm{E}[Y_{i,j}] = \sum_{0 \le i < n} \mathrm{prob}(Y_{i,j} = 1) \ .$$

We finish the proof by observing that a particular ball ends up in bin j with probability $1/m$. Thus $\mathrm{E}[X_j] = n/m$.

We conclude that the expected occupancy of any bin is n/m. That's what was to be expected, wasn't it? Go back to Table 1 with this knowledge.

For the maximal occupancy of any bin we have to work harder. We will only consider the special case that the number of balls thrown is equal to the number of bins, i.e., $n = m$. The general case can be handled similarly, but the formulae become more complicated. We defined the maximal occupancy as $Xmax = \max(X_1, \ldots, X_n)$. We can now make use of the work invested in our first proof. Recall that $p_{j,l} = \binom{n}{l}(1/n)^l(1 - 1/n)^{n-l}$ is the probability that exactly l balls end up in the j-th bin and hence $q_{j,l} = \sum_{i \ge l} p_{j,i}$ is the probability that l or more balls end up in bin j. If $Xmax \ge l$ then there must be one bin containing l or more balls and hence $\mathrm{prob}(Xmax \ge l) \le n \cdot q_{j,l}$. We also have the trivial bound $\mathrm{prob}(Xmax \ge l) \le 1$. With Lemma 4 we obtain

$$\mathrm{E}[Xmax] = \sum_{l > 0} \mathrm{prob}(Xmax \ge l) \le \sum_{l > 0} \max(1, nq_{j,l}) \ .$$

It remains to estimate the latter sum. Observe first that $p_{j,l} = \binom{n}{l}(1/n)^l(1 - 1/n)^{n-l} \le n^l/l! \cdot 1/n^l = 1/l!$ and hence

$$q_{j,l} = \sum_{i \ge l} p_{j,i} \le \sum_{i \ge l} 1/i! \le 1/l! \cdot \sum_{i \ge 0} 2^{-i} = 2/l! \ .$$

Let l_0 be the smallest integer such that $n \cdot 2/l_0! \le 1$. For $l \ge l_0$ we have $n \cdot q_{i,l} \le n \cdot 2/l! \le (n \cdot 2/l_0!) \cdot 2^{l_0 - l} \le 2^{l_0 - l}$ and hence

$$\begin{aligned} \mathrm{E}[Xmax] &\le \sum_{l > 0} \max(1, n \cdot q_{i,l}) \\ &\le \sum_{1 \le l \le l_0} 1 + \sum_{l > l_0} 2^{l_1 - l} \\ &\le l_0 + 1 \ . \end{aligned}$$

We still need an estimate for l_0: it is the smallest integer such that $l_0! \ge 2n$. We want to derive an upper bound for l_0. Since $l_0! \ge (l_0/e)^{l_0}$ (Theorem 1) we have $l_0 \le L + 1$ where L is the minimal real number with $(L/e)^L = 2n$. Taking logarithms we obtain $L \log L/e = \log 2n$ or $L = (\log 2n)/(\log L/e)$. The problem with this equation is that is has L on both sides. We use a method that is known as *bootstrapping*. We start with a fairly week bound on L and iteratively improve it. We certainly have $L \ge e$ and hence $L \le \log 2n$. Thus $L \ge (\log 2n)/(\log(\log 2n)/e) \ge (\log n)^{1/2}$, where the last inequality holds for n sufficiently large (certainly for $n \ge 2^{10}$). Substituting again, we obtain $L = (\log 2n)/(\log L/e) \le (\log 2n)/\log((\log n)^{1/2}/e) = (\log 2n)/((\log\log n)/2 - \log e) \le (4 \log 2n)/\log\log n$, where the last inequality holds for n sufficiently large (certainly $n \ge 2^{2^8}$ suffices).

We summarize our discussion in the following theorem.

Theorem 3 (Balls into Bins) *The expected occupancy of any bin after throwing n into m bins is $\beta = n/m$. For $n = m$, the expected maximal occupancy is bounded by $L + 2$ where L is minimal such that $(L/e)^L = 2n$. We have $L \leq (4\log 2n)/\log\log n$ for sufficiently large n.*

We next turn to the number of left to right maxima in random permutations. Recall from Section 0.1.1 that the number N of left-to-right maxima in a permutation $x_1x_2\ldots x_n$ of the first n integers is the number of i such that $x_i = \max(x_1, \ldots, x_i)$.

Theorem 4 (Left-to-Right Maxima) *The expected number of left-to-right maxima in a random permutation of the n distinct integers is $H_n = \sum_{1\leq i\leq n} 1/i$.*

Proof: We define the indicator random variables X_i by $X_i = 1$ is x_i is a maximum and $X_i = 0$, otherwise. Then $N = \sum_{1\leq i\leq n} X_i$ and hence by linearity of expectations and the fact that the X_i are indicator variables

$$\mathrm{E}[N] = \sum_i \mathrm{E}[X_i] = \sum_i \mathrm{prob}(X_i = 1)\ .$$

Next observe that x_i is a maximum with probability $1/i$ since there are $i!$ permutations of the first elements out of which $(i-1)!$ have x_i as a maximum. Thus

$$\mathrm{E}[N] = \sum_{1\leq i\leq n} 1/i\ .$$

∎

TODO

The number $H_n = \sum_{1\leq i\leq n} 1/i$ is called the n-th *harmonic number*. TODO: give an estimate.

It gives insight to prove the same result in different ways. Let EN_n denote the expected number of left to right maxima in a random permutation of n distinct integers. Then $EN_0 = 0$, $EN_1 = 1$ and $EN_n = 1/n + EN_{n-1}$ where the latter equality can be seen as follows. The last element is equal to n with probability $1/n$ and the first $n-1$ elements form a random permutation of $n-1$ distinct integers. From the recursion equations for EN_n one proves $EN_n = H_n$ by a simple induction.

Or we can argue as follows: $EN_n = 1+(1/n)\cdot\sum_{1\leq n} EN_{n-i}$ since the first element in a sequence is always a maximum. The first element is the i-largest with probability $1/n$ for all i. Deleting the first element and all elements smaller than it, gives us a random permutation of $n-i$ elements. Again, we can use induction to prove $EN_n = H_n$.

Exercises:

1. Estimate the maximal occupancy in the ball and bins game for $n = 100m$, $n = m\log\log n$, $n = m\log n$ and $n = m^2$.

2. Prove that $EN_0 = 0$, $EN_1 = 1$, and $EN_n = 1 + (1/n)\cdot\sum_{1\leq n} EN_{n-i}$ for $n > 1$ implies $EN_n = H_n$.

0.1.5 Chernoff Bounds

0.1.6 Conditional Probability and Independence

0.1.7 Bayes's Rule

0.1.8 Continuous Probability Spaces

Continuous probability spaces are mathematically more demanding that discrete probability spaces, because they require to deal with integrals instead of sums. In discrete probability spaces every subset of S was an event and the probability of an event was simply the sum of the probabilities of the elementary events contained in it.

In continuous spaces, only certain subsets of S are events. The family of events $\{E\}$ must satisfy the following properties: (1) the empty set is an event, (2) the complement of an event is an event, and (3) the family of events is closed under finite and countable unions and intersections.

A probability distribution prob( ) assigns a non-negative real number to each event. It must satisfy the axioms of probability theory:

- $\text{prob}(S) = 1$.
- $\text{prob}(\cup_i E_i) = \sum_i \text{prob}(E_i)$ for any finite or countable infinite family $\{E_i\}$ of pairwise mutually exclusive events.

We recommend that the readers tries to prove the five properties of probability distributions that we stated at the beginning of Section 0.1.2.

Continuous probability distributions are usually defined via density functions and integrals. We give some examples. Let I is be an interval of real numbers or the entire real line and let $f : I \mapsto I\!R_{\geq 0}$ be such that $\int_I f(x)\ \mathrm{d}x = 1$; f is called a *density function*. If $I = [a, b]$ is an interval, we can take $f(x) = 1/(b - a)$, and if I is the entire real line, we can take $f(x) = e^{x^2}/(2\pi)$. As our family of events we take all subsets $E \subseteq I$ for which the integral $\int_E f(x)\ \mathrm{d}x$ is defined. As the definition of $\text{prob}(E)$ we take the value of this integral.

Bibliography

[Knu73] D.E. Knuth. *The Art of Computer Programming (Volume I): Fundamental Algorithms*. Addison-Wesley, 1973.