Chapter 3

Inclusion-Exclusion Principle

The inclusion-exclusion principle is one of the oldest method in combinatorics, number theory discrete mathematics, and probabilistic analysis. It allows to compute either the probability that exactly r events occur out of n events (probabilistic inclusion-exclusion principle) or the number of objects that belongs exactly to r sets out of n possibly intersecting sets (combinatorial inclusion-exclusion principle). In this chapter, we derive a general form of the inclusion-exclusion principle, and discuss three applications, namely: depth in a trie, order statistics, and the longest aligned matching word.

In combinatorics we often must count the number of objects that satisfies certain properties or belongs to certain sets. The (combinatorial) principle of inclusion and exclusion is very helpful. Imagine the following situation: Sets A and B are subset of S, and we wish to count the number of objects that belong neither to A nor to B. Certainly it is not $|S|-|A|-|B|$ since the objects in $A\cap B$ have been subtracted twice. The correct answer is $|S|-|A|-|B|+|A\cap B|$, as one may expected.

In probability, we often need to compute the probability that at least one event out of n events $A_1, \ldots, A_n$ does occur. The events $A_1, \ldots, A_n$ might be dependent and not disjoint. For example, we all know that $\Pr\{A\cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A\cap B\}$. In this chapter we generalize these examples and discuss some applications of the inclusion-exclusion principle.

3.1 Probabilistic Inclusion-Exclusion Principle

Let $A_1, \ldots, A_n$ be n events which may be intersecting and dependent. We want to compute the probability P_r that exactly r events occur in a random trial. If we write $\bar{A}_i$ for the complement of A_i, then naturally

$$P_r = \sum_{1\le j_1<\cdots<j_r\le n} \Pr\{A_{j_1}\ldots A_{j_r}\bar{A}_{j_{r+1}}\ldots\bar{A}_{j_n}\} \tag{3.1}$$

where $j_1, \ldots, j_n$ is a permutation of $(1, 2, \ldots, n)$. To simplify further derivations, let us define $J_r = \{(j_1, \ldots, j_r) : 1 \le j_1 < \cdots < j_r \le n\}$ and for $k \ge 1$

$$S_k := \sum_{J_k} \Pr\{A_{j_1}\ldots A_{j_k}\} \tag{3.2}$$

with $S_0 = 1$. Then

$$\begin{aligned}\Pr\{A_{i_1}\dots A_{i_k}\} &= \Pr\{A_{i_1}\dots A_{i_k}(A_{i_{k+1}}\cap \bar{A}_{i_{k+1}})\dots(A_{i_n}\cap\bar{A}_{i_n})\} \\ &= \sum_{r=k}^{n}\sum_{J_r}\Pr\{A_{j_1}\dots A_{j_r}\bar{A}_{j_{r+1}}\dots\bar{A}_{j_n}\}\end{aligned}$$

where the second summation is over all $(j_1,\dots,j_r)$ containing $(i_1,\dots,i_k)$. Since there are $\binom{r}{k}$ combinations $(i_1,\dots,i_k)$ belonging to $(j_1,\dots,j_r)$, we finally obtain

$$S_k = \sum_{r=k}^{n}\binom{r}{k}P_r.$$

But, we are interested in P_r rather than in S_k, so we need to inverse the above. This can be done as follows:

$$\sum_{k=r}^{n}(-1)^{r+k}\binom{k}{r}S_k = \sum_{k=r}^{n}\sum_{r=k}^{n}(-1)^{r+k}\binom{k}{r}\binom{r}{k}P_r = P_r.$$

where we used the fact that $\binom{k}{r}\binom{r}{k} = 1$ if and only if $k = r$. The above equation is example of the so-called *inversion formula.*

Thus, we have just proved the *principle of inclusion and exclusion.* Observe that we did not make any assumption regarding the events $A_1,\dots,A_n$.

Theorem 3.1 (Inclusion-Exclusion Principle) *Let $A_1,\dots,A_n$ be events in an arbitrary probability space, and let P_r be the probability of exactly r of them to occur as defined precisely in (3.1). Then*

$$\begin{aligned}P_r &= \sum_{k=r}^{n}(-1)^{r+k}\binom{k}{r}\sum_{1\le j_1<\dots<j_k\le n}\Pr\{A_{j_1}\dots A_{j_k}\} \qquad (3.3)\\ &= \sum_{k=r}^{n}(-1)^{r+k}\binom{k}{r}S_k\end{aligned}$$

where S_k is explicitly defined in (3.2).

We shall discuss some applications of the inclusion-exclusion principle in the application section of this chapter. However, for the reader convenience we provide below some examples that illustrate the method.

Example 3.1: *Computing a Distribution Through Its Moments*

This example is adopted from Bollobás [31]. Let X be a random variable defined on $\{0, 1, \ldots, n\}$, and let $\mathbf{E}_k[X] = \mathbf{E}[X(X-1)\cdots(X-k+1)]$ be the kth factorial moment of X. We shall prove that

$$\Pr\{X = r\} = \frac{1}{r!}\sum_{k=r}^{n}(-1)^{r+k}\frac{\mathbf{E}_k[X]}{(k-r)!} .$$

Indeed, let $A_i = \{X \geq i\}$ for all $i = 1, \ldots, n$ where J_k is defined above. We evaluate now $S_k = \sum_{J_k} \Pr\{\bigcap_{j \in J_k} A_j\}$. Observe that

$$\begin{aligned} S_1 &= \sum_{j=1}^{n} \Pr\{X \geq j\} = \sum_{i=1}^{n} i \Pr\{X = i\} = \mathbf{E}[X], \\ S_2 &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} \Pr\{X \geq j\} = \sum_{j=2}^{n}(j-1)\Pr\{X \geq j\} \\ &= \sum_{k=2}^{n} \Pr\{X = k\}\sum_{j=1}^{k-1} j = \frac{\mathbf{E}_2[X]}{2}. \end{aligned}$$

In Exercise 1 we are asked the reader to prove the following easy generalization

$$S_k = \sum_{1 \leq j_1 < \cdots < j_k \leq n} \Pr\{X \geq j_1, \ldots, X \geq j_k\} = \frac{1}{k!}\mathbf{E}_k[X]. \tag{3.4}$$

But $\{X = r\}$ is equivalent to the event that exactly r of A_i occur, since $\{X = r\} = A_1 A_2 \ldots A_r \bar{A}_{r+1} \ldots \bar{A}_n = P_r$ (other events contributing to the probability P_r are empty). By Theorem 3.1 we prove the announced result. □

As mentioned above, we did not make any assumptions regarding A_i, and it can live on arbitrary spaces. In fact, such a generalization of the inclusion-exclusion principle is due to Fréchet. Its usefulness is illustrated in the next example.

Example 3.2: *Laplace's Formula on the Sum of I.I.D. Uniform Distributions*

Let $X_1, \ldots, X_n$ be independent random variables uniformly distributed on the interval $(0, 1)$. We prove Laplace's formula, namely (cf. [207])

$$\Pr\{X_1 + \cdots + X_n < x\} = \frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-k)_+^n$$

where $x_+ = \max\{0, x\}$. Observe that the distribution of $X_1 + \cdots + X_n$ is equal to the volume of the following set

$$\Omega = \{(x_1, \ldots, x_n) : \; x_1 + \cdots + x_n < x, \; 0 \leq x_i \leq 1, \; i = 1, \ldots, n\}.$$

To compute this we use the Fréchet formula (i.e., Theorem 3.1 applied to $\mathbb{R}^n$ space). Define the simplex

$$\widetilde{\Omega} = \{(x_1, \ldots, x_n) : \; x_1 + \cdots + x_n < x, \; x_i \geq 0, \; i = 1, \ldots, n\},$$

and let us postulate that the point $(x_1, \ldots, x_n) \in \widetilde{\Omega}$ has property A_i if $x_i > 1$. Clearly, the volume $V(\Omega)$ equal to the volume of $\widetilde{\Omega}$ minus those parts that do not belong to the unit cube $[0,1]^n$. But, this can be estimated by applying the inclusion-exclusion principle with a condition that none of A_i occurs, that is, by setting $r = 0$ in Theorem 3.1. We obtain

$$\Pr\{X_1 + \cdots + X_n < x\} = V(\Omega) = \sum_{k=0}^{n} (-1)^k S_k$$

where $S_0 = V(\widetilde{\Omega})$ and

$$S_k = \sum_{1 \leq j_1 < \cdots < j_k \leq n} V(A_{j_1}, \ldots, A_{j_k}).$$

In the above, $V(A_{j_1}, \ldots, A_{j_k})$ is the volume of the part of $\widetilde{\Omega}$ that possesses properties $A_{j_1}, \ldots, A_{j_k}$, that is, $x_{j_1} > 1, \ldots, x_{j_k} > 1$. To compute $V(A_{j_1}, \ldots, A_{j_k})$ we must set k variables in the equation $x_1 + \cdots + x_n = x$ to 1, hence

$$V(A_{j_1}, \ldots, A_{j_k}) = \frac{(x-k)_+^n}{n!}.$$

This competes the proof of Laplace's formula. □

In many situations, one needs to estimate the probability that *at least* one event occurs. This is also known as the *inclusion-exclusion principle*, and in fact it follows directly from Theorem 3.1. Indeed

$$\begin{aligned}
\Pr\{\bigcup_{i=1}^{n} A_i\} &= \sum_{r=1}^{n} P_r \\
&= \sum_{r=1}^{n} \sum_{k=r}^{n} (-1)^{r+k} \binom{k}{r} S_k = \sum_{k=1}^{n} (-1)^k S_k \sum_{r=1}^{k} \binom{k}{r} (-1)^r \\
&= \sum k = 1^n (-1)^k S_k [(1-1)^k - 1] = \sum_{k=1}^{n} (-1)^{k+1} S_k.
\end{aligned}$$

We summarize this finding in the next corollary.

Corollary 3.2 *For arbitrary events the following holds*

$$\Pr\{\bigcup_{i=1}^{n} A_i\} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J_k} \Pr\{\bigcap_{j \in J_k} A_j\} = \sum_{k=1}^{n} (-1)^{k+1} S_k. \tag{3.5}$$

where $J_k = \{1 \leq j_1 < \cdots < j_k \leq n\}$.

It is indeed quite rare to be able to compute all the probabilities involved in the inclusion-exclusion formula. Therefore, often one must retreat to inequalities. The most simple yet still very powerful is the following Boole's inequality

$$\Pr\{\bigcup_{i=1}^{n} A_i\} \leq \sum_{i=1}^{n} \Pr\{A_i\} \tag{3.6}$$

which can be easily proved by induction. However, we can do much better. The reader is asked in Exercise2 to use mathematical induction to prove the following Bonferroni's inequalities.

Lemma 3.3 (Bonferroni's Inequalities) *For every even integer $m \geq 0$ we have*

$$\begin{aligned}\sum_{j=1}^{m}(-1)^{j-1} \sum_{1\leq t_1<\cdots<t_j\leq n} \Pr\{A_{t_1}\ldots\mathcal{A}_{t_j}\} &\leq \Pr\{\bigcup_{i=1}^{n} A_i\}\\ &\leq \sum_{j=1}^{m+1}(-1)^{j-1} \sum_{1\leq t_1<\cdots<t_j\leq n} \Pr\{A_{t_1}\ldots A_{t_j}\} .\end{aligned}$$

In Section 3.3.3 below, we use Bonferroni's inequalities to derive the limiting distribution of the longest aligned match between two sequences.

3.2 Combinatorial Inclusion-Exclusion Principle

In this section, we count the number of objects (properties) that belong exactly to r (possibly intersecting) subsets. Imagine the following scenario: There are N objects in a set X whose subsets are $A_1, \ldots, A_n$ of cardinalities $N_1, \ldots, N_n$, respectively. We are interested in counting the number, $N(r)$, of objects that belong exactly to r subsets, where $r = 0, 1, \ldots, n$. Let $N_{j_1,\ldots,j_k} = |A_{j_1} \cap \cdots \cap A_{j_k}|$, that is, $N_{j_1,\ldots,j_k}$ is the cardinality of the intersection of the subsets $A_{j_1}, \ldots, A_{j_k}$ where $1 \leq j_1 < \cdots < j_k \leq n$. The following, also known as the *combinatorial inclusion-exclusion principle*, is a direct consequence of Theorem 3.1.

Theorem 3.4 *For any n and $0 \leq r \leq n$, the number $N(r)$ of objects that belong to exactly r subsets $A_1, \ldots, A_n$ of a set X of cardinality N becomes*

$$\begin{aligned} N(r) &= \sum_{k=r}^{n}(-1)^{r+k}\binom{k}{r} \sum_{1\leq j_1<\cdots<j_k\leq n} N_{j_1,\ldots,j_k} \qquad (3.7)\\ &= \sum_{k=r}^{n}(-1)^{r+k}\binom{k}{r} W_k \end{aligned}$$

where $W_k = \sum_{1\leq j_1<\cdots<j_k\leq n} N_{j_1,\ldots,j_k}$.

Proof. We derived the above directly from the probabilistic inclusion-exclusion principle. It suffices to assume that all objects in X are uniformly distributed and define the probability of a subset A of X as $\Pr\{A\} = |A|/N$. Then, the probability P_r of Theorem 3.1 becomes $N(r)/N$, and (3.7) follows immediately from (3.3). ■

We illustrate the combinatorial inclusion-exclusion principle on a few examples. More can be found in the exercises at the end of this chapter. We start with a simple enumeration of surjections.

Example 3.3: *Number of Surjections*

Let X be a set of cardinality n, and $Y = \{y_1, \ldots, y_m\}$. We want to enumerate the number of functions (surjections) from X on Y. Let us call this number s_{nm}. Clearly, $s_{nm} = n!$ if $n = m$ and $s_{nm} = 0$ if $m > n$. This can be used to verify the formula that we are going to derive. Let A_i be a subset of all surjections $f : X \to Y$ such that $y_i \notin f(X)$. In terms of Theorem 3.4 we need to find $N(0) = s_{nm}$, that is, the number of functions that *do not* belong to any of A_i. But $|A_{j_1} \cap \cdots \cap A_{j_k}| = (m-k)^n$, thus by (3.7) with $r = 0$ we obtain

$$s_{nm} = \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n.$$

Interestingly enough, we also proved that

$$\begin{aligned} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^n &= n! \\ \sum_{k=0}^{m} (-1)^k \binom{m}{k} (m-k)^n &= 0 \quad \text{for} \quad m > n \end{aligned}$$

Such combinatorial relationships are sometimes quite cumbersome to prove. Counting can help! □

Let us now use Theorem 3.4 to count sequences satisfying certain properties. In those problems the so-called *Möbius function* plays a pivotal role. It is defined as follows: consider an integer n that has a unique factorization as product of prime powers, that is,

$$n = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r} \tag{3.8}$$

where p_i are distinct prime numbers and all the powers $e_i \geq 1$. We define the Möbius function $\mu(n)$ as $\mu(1) = 1$ and for $n > 1$

$$\mu(n) = \begin{cases} 0 & \text{if any power } e_i > 1 \\ (-1)^r & \text{if } e_1 = \cdots = e_r = 1 \end{cases} \tag{3.9}$$

The Möbius function finds applications in counting problems that involve divisors of n. Let us agree to write $d|n$ when d is a divisor of n. Observe that

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

where the summation is over all divisors d of n. Indeed, to prove the above, let n has the factorization as in (3.8). By the definition of the Möbius function we have

$$\sum_{d|n} \mu(d) = \sum_{i=0}^{r} \binom{r}{i} (-1)^i = (1-1)^r = 0$$

since one finds $\binom{r}{i}$ divisors d consisting of i distinct primes involved in the factorization of n.

This leads us to the following interesting *inversion formula.*

Theorem 3.5 (Möbius Inversion Formula) *Let $f(n)$ and $g(n)$ be functions defined for positive integer n that satisfy*

$$f(n) = \sum_{d|n} g(d).$$

Then

$$g(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) \tag{3.11}$$

Proof. By (3.10) we have

$$\begin{aligned}
\sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) &= \sum_{d|n} \mu\left(\frac{n}{d}\right) f(d) \\
&= \sum_{d|n} \mu\left(\frac{n}{d}\right) \sum_{d'|d} g(d') = \sum_{d'|n} g(d') \sum_{m|(n/d')} \mu(m) \\
&= g(n)
\end{aligned}$$

as desired. ∎

The inversion formula like the above is very useful in counting and proving combinatorial identities. We shall return to them in Chapter 7 where we use analytic tools to explore them (cf. Section 7.4.2). We finish this section with an application of Theorem 3.5.

Example 3.4 *Primitive Sequences*

Let us count the number of binary sequences of length n that are primitive, that is, they are *not* expressible as a concatenation of some identical smaller sequences. For example, 010101 is not a primitive sequence while 011010 is a primitive one. Let $f(n)$ the number of such primitive sequences. But there are 2^n all binary strings, and every one is uniquely expressible as a concatenation of n/d identical *primitive* strings of length d, where d is a divisor of n (e.g., $(01)(01)(01)$ while (011010)). Thus for every n

$$2^n = \sum_{d|n} f(d).$$

Then, by the Möbius inversion formula we have

$$f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) 2^d$$

for the required number. □

3.3 Applications

We now tackle some problems in which the inclusion-exclusion principle play a significant role. We start with deriving the generating function for the depth in a trie, then deal with order statistics, and finish with deriving the asymptotic distribution for the longest aligned match words in two randomly generated strings. The last problem is by far the most challenging.

3.3.1 Depth in a Trie

In Chapter 1.1 we discussed tries built over n binary strings $X^1, \ldots, X^n$. We assume that those strings are generated by a binary memoryless source with "0" occurring with probability p and '1" with probability $q = 1 - p$. We are interested here in the depth $D_n(m)$ as defined in Definition 1.1. By Definition 1.2, C_{ij} is the length of the longest string that is a prefix of X^i and X^j. Then, by Theorem 1.3

$$D_n(m) = \max_{1 \leq i \neq m \leq n} \{C_{i,m}\} + 1. \tag{3.12}$$

Certainly, the alignments C_{ij} are dependent random variables even for the memoryless source.

Our goal is to compute the generating function of the depth $D_n := D_n(1)$, that is, $D_n(u) = \sum_{k=0}^{\infty} \Pr\{D_n = k\} u^k = \mathbf{E}[u^{D_n}]$. Observe that

$$\Pr\{D_n > k\} = \Pr\{\bigcup_{i=2}^{n} A_{i,1}\}$$

where $A_{ij} = \{C_{ij} \geq k\}$. Thus, one can apply the principle of inclusion-exclusion as expressed in Corollary 3.2 to obtain

$$\Pr\{D_n > k\} \quad = \quad \Pr\{\bigcup_{i=2}^{n} [C_{i,1} \geq k]\}$$

$$= \sum_{r=1}^{n-1}(-1)^{r+1}\binom{n-1}{r}\Pr\{C_{2,1} \geq k, \ldots, C_{r+1,1} \geq k\}$$

since the probability $\Pr\{C_{2,1} \geq k, \ldots, C_{r+1,1} \geq k\}$ does not depend on the choice of strings (i.e., it is the same for any r-tuple of selected strings). Moreover, it can be easily explicitly computed. Indeed, we find

$$\Pr\{C_{2,1} \geq k, \ldots, C_{r+1,1} \geq k\} = (p^{r+1} + q^{r+1})^k$$

since r independent strings must agree on the first k symbols, and the probability they agree on a symbol at a given position is $p^{r+1} + q^{r+1}$. Thus,

$$\sum_{k=0}^{n} \Pr\{D_n > k\}u^k = \sum_{r=1}^{n-1}(-1)^{r+1}\binom{n-1}{r}\frac{1}{1 - u(p^{r+1} + q^{r+1})}.$$

In Exercise 4 we asked the reader to prove that

$$\sum_{k=0}^{\infty} \Pr\{D_n > k\}u^k = \frac{1 - \mathbf{E}[u^{D_n}]}{1 - u}$$

which finally leads to

$$D_n(u) = 1 + \sum_{r=1}^{n-1}(-1)^{r}\binom{n-1}{r}\frac{1-u}{1 - u(p^{r+1} + q^{r+1})}$$

for $|u| < 1$.

3.3.2 Order Statistics

Let $Y_1, \ldots, Y_m$ be a sequence of random variables. We often need to compute the maximum or minimum values of such a sequence. In some problems (e.g., combinatorial optimization problems as discussed in Chapter 1.6) we are also interested in the r-th greatest value of $Y_1, \ldots, Y_m$, where $r = 1, \ldots, m$. More precisely, let $\min_{1 \leq i \leq m}\{Y_i\} := Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(m)} := \max_{1 \leq i \leq m}\{Y_i\}$. We call $Y_{(r)}$ the rth *order statistic* of $Y_1, \ldots, Y_m$.

The inclusion-exclusion principle can be used to find probabilistic behavior of order statistics. Indeed, let $A_i = \{Y_i > x\}$, then $\{Y_{(m)} > x\} = \{\max\{Y_1, \ldots, Y_m\} > x\} = \bigcup_{i=1}^m A_i$. More interestingly, we can express the rth order statistic as follows

$$\{Y_{(r)} > x\} = \bigcup_{1 \leq j_1 < \cdots < j_{m-r+1} \leq m} \ \bigcap_{i=1}^{m-r+1} A_{j_i} \tag{3.13}$$

since for $\{Y_{(r)} > x\}$ to hold it is required that at least $m-r+1$ variables out of $Y_1, \ldots, Y_m$ are greater than x. Indeed, for $r = m$ we have $\{Y_{(m)} > x\} = \{\max_{1\le i\le m}\{Y_i\} > x\} = \bigcup_{i=1}^m \{Y_i > x\}$, and for $r = m-1$ we obtain

$$\begin{aligned}\{Y_{(m-1)} > x\} &= \{Y_1 > x \;\&\; Y_2 > x \text{ or} \ldots \text{or } Y_1 > x \;\&\; Y_m > x \\ &\quad \text{or} \ldots \text{or } Y_{m-1} > x \;\&\; Y_m > x\}\end{aligned}$$

and the general case is given in (3.13).

The next finding presents a simple result on the maximum of a set of dependent and independent random variables.

Theorem 3.6 *Let Y_1, $Y_2, \ldots, Y_m$ be a sequence of random variables with distribution functions $F_1(y)$, $F_2(y)$,..., $F_m(y)$, respectively, and let $R_i(y) = \Pr\{Y_i \ge y\}$. Define $Y_{(m)} = \max_{1\le i\le m}\{Y_i\}$.*

(i) *If a_m is the smallest solution of*

$$\sum_{k=1}^{m} R_k(a_m) = 1 \ , \tag{3.14}$$

then

$$\mathbf{E}[Y_{(m)}] \le a_m + \sum_{k=1}^{m} \sum_{j=a_m}^{\infty} R_k(j) \ .$$

(ii) *If the distribution function $F_i(y)$ of Y_i satisfies for all $1 \le i \le m$ the following two conditions*

$$F_i(y) < 1 \qquad \text{for all} \quad y < \infty \ ,$$

and

$$\lim_{y\to\infty} \sup_i \frac{1 - F_i(cy)}{1 - F_i(y)} = 0 \qquad \text{for} \quad c > 1 \ , \tag{3.15}$$

then $Y_{(m)}/a_m \le 1$ in probability (pr.), that is, for any $\varepsilon > 0$

$$\lim_{m\to\infty} \Pr\{Y_{(m)} \ge (1+\varepsilon)a_m\} = 0 \ ,$$

where a_m solves asymptotically (3.14).

(iii) *If $Y_1, \ldots, Y_m$ are independently and identically distributed with common distribution function $F(\cdot)$, then $Y_{(m)} \sim a_m$ (pr.), that is, for any $\varepsilon > 0$*

$$\lim_{m\to\infty} \Pr\{(1-\varepsilon)a_m \le Y_{(m)} \le (1+\varepsilon)a_m\} = 1 \ ,$$

where a_m is a solution of (3.14) which in this case becomes $mR(a_m) = 1$.

Proof. For (i) observe that for any a

$$Y_{(m)} \leq a + \sum_{k=1}^{m} [Y_k - a]^+$$

where $t^+ = \max\{0,\ t\}$. Since $[Y_k - a]^+$ is a nonnegative random variable, then $\mathbf{E}[Y_k - a]^+ = \int_a^\infty R_k(y)dy$, (assuming for simplicity that Y_k is a continuous random variable; cf. Exercise 1), thus

$$\mathbf{E}[Y_{(m)}] \leq a + \sum_{k=1}^{m} \int_a^\infty R_k(x)dx\ .$$

Minimizing the right-hand side of the above with respect to a yields part (i). For part (ii) we use (3.13) for $r = m$, and after a simple application of (3.6) we arrive at

$$\Pr\{Y_{(m)} > x\} = \Pr\{\bigcup_{i=1}^{m} Y_i > x\} \leq \sum_{i=1}^{m} R_i(x).$$

Let now $x = (1+\varepsilon)a_m$ where $\varepsilon > 0$ and a_m be defined by (3.14). Then, (3.15) with $c = 1+\varepsilon$ implies $R((1+\varepsilon)a_m) = o(1)R(a_m)$, hence

$$\Pr\{Y_{(m)} \geq (1+\varepsilon)a_m\} = o(1)\sum_{i=1}^{m} R(a_m) \to 0.$$

Part (iii) can be proved using the additional assumption about the independence of $Y_1, \ldots, Y_m$. The reader is asked in Exercise 9 to provide details of the proof. ∎

Finally, we present a simple asymptotic result concerning the behavior of $Y_{(r)}$ for the so-called exchangeable random variables. A sequence $\{Y_i\}_{i=1}^m$ is *exchangeable* if for any k-tuple $\{j_1, \ldots, j_k\} \subset \{1, \ldots, m\}$ the following holds

$$\Pr\{Y_{j_1} < y_{j_1}, \ldots, Y_{j_k} < y_{j_k}\} = \Pr\{Y_1 < y_1, \ldots, Y_k < y_k\},$$

that is, the joint distribution depends only on the *number* of variables (cf. [94]).

Theorem 3.7 *Let $\{Y_i\}_{i=1}^m$ be exchangeable random variables, and let $R_r(x) = \Pr\{Y_1 > x, \ldots, Y_r > x\}$ satisfy for all $c > 1$*

$$\lim_{x\to\infty} \frac{R_{m-r+1}(cx)}{R_{m-r+1}(x)} = 0 \qquad (3.16)$$

uniformly over all $r = 1, \ldots, m$. Define $a_m^{(r)}$ as the smallest solution of

$$\binom{m}{r-1} R_{m-r+1}(a_m^{(r)}) = 1.$$

Then asymptotically $Y_{(r)} \leq a_m^{(r)}$ (pr.).

Proof. We use (3.13) with $x = (1+\varepsilon)a_m^{(r)}$. By (3.6) and the assumption of exchangebility, we obtain

$$\begin{aligned}\Pr\{Y_{(r)} > (1+\varepsilon)a_m^{(r)}\} &\le \binom{m}{n-r+1} R_{m-r+1}((1+\epsilon)a_m^{(r)}) \\ &= \binom{m}{r-1} R_{m-r+1}(a_m^{(r)})o(1) \to 0,\end{aligned}$$

where the last equation follows from (3.16). ∎

The reader is asked in Exercise 9–11 to extend the above results and apply them to some combinatorial optimization problems.

3.3.3 Longest Aligned Matching Words

We report here some results of Karlin and Ost [136]. Imagine two sequences of length n, say $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$. We assume that both are generated by memoryless sources over a binary alphabet, say Bernoulli(p_1, p_2) and Bernoulli(r_1, r_2), respectively. We are interested in the length W_n of the *longest aligned match word*, that is, a word as long as possible that is common to both strings and starts at the same position in both strings. More precisely:

$$W_n := \max\{k : \exists_{1\le t\le n}\ X_t = Y_t, \ldots, X_{t+k-1} = Y_{t+k-1} \text{ and } X_{t+k} \ne Y_{t+k}\}.$$

To see how this problem is related to the topic of this section, let us observe that $W_n = \max_k\{A(n,k)\}$ where

$$A(n,k) = \{\text{there exists an aligned match word of length } k \text{ in both sequences}\}.$$

Then $\Pr\{W_n \ge k\} = \Pr\{A(n,k)\}$ and

$$A(n,k) = \bigcup_{t=1}^{n} A_t(k)$$

where $A_t(k)$ is the event signifying a k-word match starting at position $1 \le t \le n$. Thus, the principle of inclusion–exclusion (cf. Corollary 3.2) can be used to evaluate $\Pr\{A(n,k)\}$, and hence $\Pr\{W_n \ge k\}$.

We start with the formulation of the result of Karlin and Ost [136]. Let $P = p_1r_1 + p_2r_2$ be the probability of a match at a given position. Thus, the probability of a match of length at least k starting at any position is equal to P^k. For convenience we also define $Q = P^{-1}$. We prove below that the maximal aligned match W_n is concentrated around

$$k = \frac{\ln n}{\ln(P^{-1})} + x + 1 = \log_Q n + x + 1$$

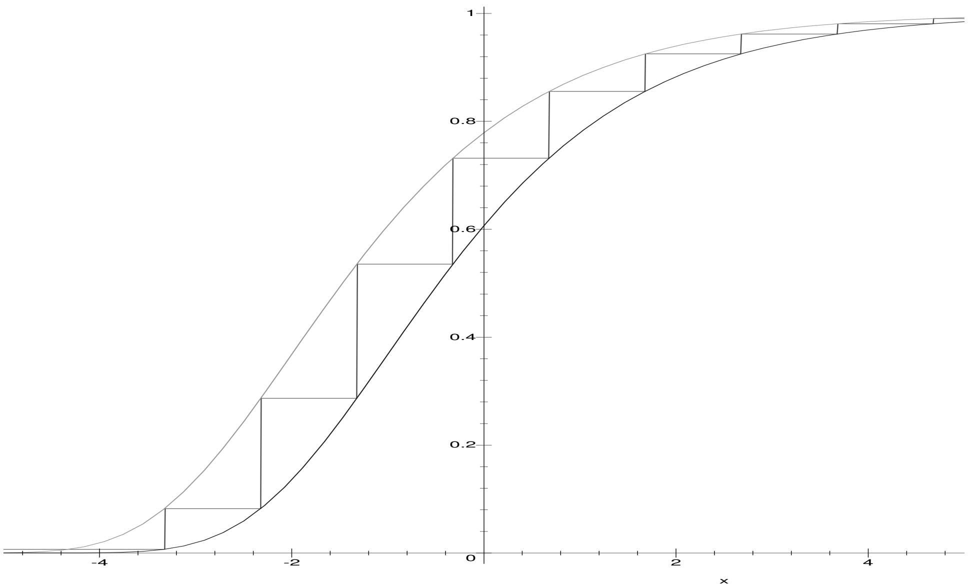

Figure 3.1: The distribution function of $W_n - \log_2 20$ with two continuous extremes for binary symmetric Bernoulli models.

where x is a real number. We denote by $\widetilde{x}$ such x (that may depend on n) that $k = \log_Q n + \widetilde{x} + 1$ is an integer.

Theorem 3.8 (Karlin and Ost 1987) *Under the above assumptions*

$$\lim_{n\to\infty} \Pr\{W_n \le \log_Q n + \widetilde{x}\} = \exp\left(-(1-P)P^{\widetilde{x}+1}\right). \tag{3.17}$$

If x is a fixed real number, then the limiting distribution of W_n does not exit, but one can claim

$$\liminf_{n\to\infty} \Pr\{W_n \le \log_Q n + x\} = \exp\left(-(1-P)P^{x}\right) \tag{3.18}$$

$$\limsup_{n\to\infty} \Pr\{W_n \le \log_Q n + x\} = \exp\left(-(1-P)P^{x+1}\right) \tag{3.19}$$

where $Q^{-1} = P = p_1 r_1 + p_2 r_2$ is the probability of a match at a given position.

Before we proceed with the proof, some explanations are in order. The limiting distribution of W_n does not exist for general real x, but an asymptotic distribution may be found which resembles the so called *extreme distribution* $\exp(-e^{-x})$ (cf. Galambos [94]). Indeed, we shall prove below that for large n

$$\lim_{n\to\infty} \Pr\{W_n \le \log_Q n + x\} \sim \exp\left(-(1-P)P^{x+1-\langle \log_Q n + x\rangle}\right)$$

where $\langle \log n\rangle = \log n - \lfloor \log n \rfloor$. This fractional part $\langle \log_Q n\rangle$ behaves very erratically. We shall study in depth its properties in Section 8.2.3. The asymptotic distribution is presented in Figure 3.1 where the staircase like function $\exp\left(-(1-P)P^{x+1-\langle \log_Q n + x\rangle}\right)$ is plotted together with the lower extreme distribution $\exp(-(1-P)P^x)$ and the upper extreme distribution $\exp(-(1-P)P^{x+1})$. We should point out, however, that such a behavior is not a surprise since W_n as a discrete order statistic is expected not to have a limiting distribution. Already in 1970 Anderson [8] observed that the limiting distribution of maximum of *discrete* i.i.d. random variables may not exist.

Now we are ready to prove the Karlin and Ost result. Throughout this derivation we assume that

$$k = \frac{\ln n}{\ln(P^{-1})} + \widetilde{x} + 1 = \log_Q n + \widetilde{x} + 1 \tag{3.20}$$

where $\widetilde{x}$ is such that k is an integer. As observed above, we need to estimate $\Pr\{A(n,k)\}$, and after applying Corollary 3.2 we can write

$$\Pr\{A(n,k)\} = \sum_{m=1}^{n}(-1)^{m+1}\sum_{1\le t_1<\ldots<t_m\le n}\Pr\{A_{t_1}\ldots A_{t_m}\} = \sum_{s=1}^{n}(-1)^{m+1}S_m(n).$$

where $S_m(n) = \sum_{1\le t_1<\ldots<t_m\le n}\Pr\{A_{t_1}\ldots A_{t_m}\}$, and we write A_t for $A_t(k)$.

The plan for establishing the limit of $\Pr\{A(n,k)\}$ is as follows: We shall prove that for k as in (3.20) every term of the above sum converges to a function, say $a_m(\widetilde{x})$ as $n\to\infty$, that is, for $m = 1, 2, \ldots$

$$\lim_{n\to\infty} S_m(n) = \lim_{n\to\infty}\sum_{1\le t_1<\ldots<t_m\le n}\Pr\{A_{t_1}\ldots A_{t_m}\} = a_m(\widetilde{x}). \tag{3.21}$$

By Bonferroni's inequalities for every even m

$$\begin{aligned}\liminf_{n\to\infty}\Pr\{A(n,k)\} &\ge \sum_{i=1}^{m}(-1)^{i+1}a_i(\widetilde{x}),\\ \limsup_{n\to\infty}\Pr\{A(n,k)\} &\le \sum_{i=1}^{m+1}(-1)^{i+1}a_i(\widetilde{x})\end{aligned}$$

Next, we shall show that the alternating series

$$\sum_{i=1}^{\infty}(-1)^{i+1}a_i(\widetilde{x}) = a(\widetilde{x})$$

converges uniformly for bounded x, which would imply that

$$\lim_{n\to\infty}\Pr\{W_n \le \log_Q n + \widetilde{x}\} = 1 - a(\widetilde{x}).$$

We now compute $\lim_{n\to\infty} S_m(n)$ for every m. Let us start with $m = 1$. Observe that $\Pr\{A_t\} = P^k$, thus

$$\lim_{n\to\infty} S_1(n) = \lim_{n\to\infty} nP^{\log_Q n+\widetilde{x}+1} = P^{\widetilde{x}+1} := \phi(x).$$

This was quite easy! The next step requires a little more work. Let

$$\begin{aligned} S_2(n) &= \sum_{t_1<t_2} \Pr\{A_{t_1}A_{t_2}\} = \frac{1}{2}\sum_{|t_2-t_1|\le k} \Pr\{A_{t_1}A_{t_2}\} + \frac{1}{2}\sum_{|t_2-t_1|>k} \Pr\{A_{t_1}A_{t_2}\} \\ &:= T_1(n) + T_2(n). \end{aligned}$$

By independence $T_2(n)$ can be easily computed, and we arrive at

$$\lim_{n\to\infty} T_2(n) = \frac{1}{2}\lim_{n\to\infty} (S_1(n))^2 = \frac{1}{2}\phi^2(\widetilde{x}).$$

To assess $T_1(n)$, we must take into account overlapping parts. Let the two overlapping words of length k be such that the first word starts at position t and the overlapping begins at position $t+j$ for $1 \le j \le k$. Let C_2 denote the extended match of length $k+j$. Then $\Pr\{C_2 > k+j\} = P^{k+j}$, and hence

$$T_1(n) = \sum_{t=1}^{n}\sum_{j=1}^{k} P^{k+j} = nP^k\left(\frac{P}{1-P}\right)(1-P^k).$$

In summary, for k as in (3.20)

$$\lim_{n\to\infty} S_2(n) = \frac{1}{2}\left(\phi(\widetilde{x})\right)^2 + \left(\frac{P}{1-P}\right)\phi(\widetilde{x})$$

since $(1-P^k) = (1-P^{\widetilde{x}+1}/n) \to 1$ as $n\to\infty$.

Let us now consider a general m and compute $S_m(n)$. We partition $m = n_1 + 2n_2 + 3n_3 + \cdots + mn_m$ so that $S_m^{(n_1,n_2,\ldots,n_m)}(n)$ denotes the cumulation of probabilities $\Pr\{A_{t_1},\ldots,A_{t_m}\}$ of joint events which consist of n_i matches of non-overlapping k-words, n_2 pairs of overlapping k-word matches which are mutually independent, n_3 groups of overlapping three k-word matching which are disjoint between the different triples, etc., and n_m (necessarily at most 1) successive overlapping k-word matches. Then, $S_m(n) = \sum_{\mathcal{P}(n_1,\ldots,n_m)} S_m^{(n_1,n_2,\ldots,n_m)}(n)$ where $\mathcal{P}(n_1,\ldots,n_m)$ is the set of all integer partition such that $m = n_1 + 2n_2 + 3n_3 + \cdots + mn_m$. It is not too difficult to observe that

$$S_m^{(n_1,n_2,\ldots,n_m)}(n) = \frac{1}{n_1!n_2!\cdots n_m!}\prod_{i=1}^{m}\left(nP^k\left(\frac{P}{1-P}\right)^{m-1}(1-P^k)^{m-1}\right)^{n_i}.$$

Indeed, consider n_i groups of overlapping i word matches of length k. Since any permutation of these words leads to the same probability, we see the term $1/n_i!$ in the above. Let a group of i k-word matches starts at position $1 \le t \le n$ and the overlaps begin at positions $t + j_1, \ldots, t + j_m$, respectively. Such an extended C_m match has the following probability

$$\Pr\{C_m > k + j_1 + \cdots j_m\} = \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} P^{k+j_1+\cdots+j_m} = P^k \left(\frac{P}{1-P}\right)^{m-1} (1-P^k)^{m-1}$$

Summing over all $1 \le t \le n$, and noting that the n_i groups are independent, we obtain the desired formula. Thus, for k as in (3.20) we arrive at

$$\lim_{n\to\infty} S_m^{(n_1,n_2,\ldots,n_m)}(n) = \frac{1}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{m} s_i^{n_i},$$

where $s_i = \phi(\tilde{x})(P/(1-P))^{i-1} = P^{\tilde{x}+1}(P/(1-P))^{i-1}$ and finally

$$a_m(\tilde{x}) = \lim_{n\to\infty} S_m(n) = \sum_{n_1+2n_2+\cdots mn_m=m} \frac{1}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{m} s_i^{n_i}. \tag{3.22}$$

To complete the proof, we need to sun up the alternating series $\sum_{m=1}^{\infty}(-1)^{m+1}a_m(\tilde{x})$. For this we need one result from generating functions, namely

$$\sum_{m=0}^{\infty} \sum_{n_1+2n_2+\cdots mn_m=m} \frac{x^m}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{m} c_i^{n_i} = \exp\left(\sum_{i=1}^{\infty} c_i x^i\right) \tag{3.23}$$

provided the series $\sum_{i=1}^{\infty} c_i x^i$ converges. We prove it in Example 5 of Chapter 7. (This is the point where analytic and probabilistic analyses meet and need each others to solve the problem at hand.) The above formula is a simple application of the convolution formula of exponential generating functions, and will be discussed in Chapter 7. Knowing this, we can proceed as follows

$$\begin{aligned}
\sum_{m=1}^{\infty}(-1)^{m+1}a_m(\tilde{x}) &= \sum_{m=1}^{\infty}(-1)^{m+1} \sum_{n_1+2n_2+\cdots mn_m=m} \frac{1}{n_1!n_2!\cdots n_m!} \left(\frac{P}{1-P}\right)^{\sum_{i=1}^{m} n_i(i-1)} \\
&\times \ (\phi(\tilde{x}))^{\sum_{i=1}^{m} n_i} = \sum_{m=1}^{\infty}(-1)^{m+1}\left(\frac{P}{1-P}\right)^m \\
&\times \sum_{n_1+2n_2+\cdots mn_m=m} \frac{1}{n_1!n_2!\cdots n_m!} \left(\frac{1-P}{P}\phi(\tilde{x})\right)^{\sum_{i=1}^{m} n_i} \\
&= 1 - \exp\left(-(1-P)P^{\tilde{x}+1}\right)
\end{aligned}$$

where the last equality comes from identity (3.23) with $x = -P/(1-P)$ and $c_i = (1-P)\phi(\tilde{x})/P$. This proves (3.17) of Theorem 3.8. To establish the second part of Theorem 3.8, we observe that for any real x we can write $\Pr\{W_n \geq \log_Q + x + 1\} = \Pr\{W_n \geq \lfloor \log_Q + x + 1 \rfloor\}$ and $\lfloor \log_Q + x + 1 \rfloor = \log_Q + x + 1 + \langle \log_Q + x \rangle$.

3.4 Extensions and Exercises

3.1 Prove (3.4).

3.2 Prove Lemma 3.3 using mathematical induction.

3.3 (Bell's Inequality) Prove the following useful inequality of Bell

$$\Pr\{A \cap B\} \leq \Pr\{A \cap C\} + \Pr\{B \cap \bar{C}\}$$

where A, B, C are events and $\bar{C}$ is the complementary event to C. The above inequality is used in quantum physics to establish Einstein–Podolsky–Rosen paradox.

3.4 Prove that for any random variable X over nonnegative integers

$$\sum_{k=0}^{n} \Pr\{X > k\} u^k = \frac{1 - \mathbf{E}[u^X]}{1-u}$$

for $|u| < 1$.

3.5 Prove that if X is a non-negative random variable with finite moments and for all m

$$\lim_{k \to \infty} \frac{\mathbf{E}_k[X] k^m}{k!} = 0,$$

then

$$\Pr\{X = r\} = \frac{1}{r!} \sum_{k=r}^{\infty} (-1)^{r+k} \frac{\mathbf{E}_k[X]}{(k-r)!}$$

where, as before, $\mathbf{E}_k[X]$ is the kth factorial moment of X.

3.6 Enumerate the number of binary circular sequences, where two sequences obtained by a rotation are considered the same.

3.7 Let $n = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}$ be a positive integer, and denote by $\phi(n)$ the number of integers $1 \leq k \leq n$ such that $\gcd(n, k) = 1$. (The function ϕ is known as the Euler function.) Prove that

$$\phi(n) = n \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq j_1 < \ldots, j_k \leq n} \frac{1}{p_{j_1} \cdots p_{j_k}} = n \prod_{i=1} \left(1 - \frac{1}{p_i}\right),$$

and

$$\sum_{d|n} \phi(d) = n.$$

Conclude that

$$\phi(n) = n \sum_{d|n} \frac{\mu(d)}{d}$$

for all positive integers n.

3.8 How many positive integers less than 1000 have no factor between 1 and 10.

3.9 Prove part (iii) of Theorem 3.6. Then establish the following refinement: Let $Y_1, \ldots, Y_m$ be i.i.d. and let a_m be the smallest solution of $m(1 - \Pr\{Y_1 > a_m\}) = 1$. Then

$$\lim_{m\to\infty} \Pr\{\max\{Y_1, \ldots, Y_m\} = \lfloor a_m \rfloor \text{ or } \lfloor a_m \rfloor + 1\} = 1.$$

Hint. Verify that

$$\lim_{m\to\infty} (\Pr\{\max\{Y_1, \ldots, Y_m\} < x\} - \exp(n(1 - \Pr\{Y_1 > x\}))) = 0$$

for real x.

3.10 (Aldous *et al.* 1992) Prove the following result: *If $\{X_k\}_{k=1}^{\infty}$ are i.i.d. Poisson distributed random variables with mean ρ, then for large enough integers a and n*

$$\Pr\{\max_{1\le k\le n} X_k < a\} - \exp(-ne^{-\rho}\rho^a/a!) \to 0 \quad \textit{as} \quad n, a \to \infty$$

and

$$\Pr\{\max_{1\le k\le n} X_k = \lfloor a_n \rfloor + 1 \textit{ or } \lfloor a_n \rfloor\} = 1 - O(1/a_n) \to 1 \qquad \textit{as } n \to \infty.$$

For large n the sequence $\{a_n\}$ satisfies

$$a_n \sim \frac{\log n - \rho}{\log(\log n - \rho) - \log \rho} \sim \frac{\log n}{\log\log n},$$

where a_n is defined as the smallest solution of the equation

$$n \cdot \frac{\gamma(a_n, \rho)}{\Gamma(a_n)} = 1.$$

In the above $\gamma(x, \rho) \equiv \int_0^\rho t^{x-1} e^{-t} dt$ is the incomplete gamma function, and $\Gamma(x) = \gamma(x, \infty)$ is the gamma function discussed in Chapter 2.4.

3.11 Let $W = \{W_{i,j}\}_{i,j=1}^{n}$ be a matrix of *weights* that are i.i.d. distributed according to a common strictly continuous and increasing distribution $F(\cdot)$. Consider the following *bottleneck assignment problem*

$$Z_{\min} = \min_{\sigma \in B_n} \{ \max_{1 \le i \le n} w_{i,\sigma(i)} \},$$

and *capacity assignment problem*

$$V_{\max} = \max_{\sigma \in B_n} \{ \min_{1 \le i \le n} w_{i,\sigma(i)} \},$$

where B_n is a set all permutations of $\{1, 2, ..., n\}$. Prove that

$$\begin{aligned} \lim_{n\to\infty} \frac{Z_{(d)}}{F^{-1}(\log n/n)} &= 1 \qquad \text{(pr.)} \\ \lim_{n\to\infty} \frac{V_{(d)}}{F^{-1}(1-\log n/n)} &= 1 \qquad \text{(pr.)} \end{aligned}$$

where $F^{-1}(\cdot)$ denotes the inverse function of the distribution $F(\cdot)$.

3.12 (Karlin and Ost [136]) Extend Theorem 3.8 to Markov model. More generally, compute the asymptotic distribution of the longest aligned word match found in r or more sequences among s sequences generated by Markov sources.

3.13 (Karlin and Ost [137]) Compute the limiting distribution of the longest word (not necessary aligned) common to r or more of s sequences generated by memoryless and Markov sources.

Inclusion-Exclusion