0

A few things you need to know

0.1 Probability notation and prerequisites

The book assumes knowledge of the basic concepts of probability theory at the level of a first graduate course. For readers' convenience, we recall here a few standard definitions and notational conventions: first, throughout the book we use the following notation and abbreviations.

$\mathbb{P}(\cdot)$	Probability of an event
$\mathbb{E}(\cdot)$	Expectation of a random variable
$\mathbf{1}_{\{\cdot\}}$	The indicator (a.k.a. characteristic function) of an event/set
r.v.	random variable
i.i.d.	independent and identically distributed
a.s.	almost surely
$\overset{d}{=}$	equality in distribution
$\sim$	[a random variable] is distributed as [a distribution] (see below for examples)

Second, we make occasional use of the standard terminology regarding modes of convergence for sequences of random variables and probability distributions, which are defined as follows.

Almost sure convergence. We say that a sequence $(X_n)_{n=1}^\infty$ of random variables converges almost surely to a limiting random variable X, and denote $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X$, if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

Convergence in probability. We say that X_n converges in probability to X, and denote $X_n \xrightarrow[n\to\infty]{P} X$, if for any $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

In a few places, the term "convergence in probability" is used in a broader sense that applies to convergence of random objects taking value in a more general space than the real line. In such cases, the meaning of the convergence statement is spelled out explicitly.

Convergence in distribution. We say that a sequence of distribution functions F_n converges in distribution to a limiting distribution function F, and denote $F_n \xrightarrow[n\to\infty]{d} F$, if $F_n(x) \to F(x)$ for any $x \in \mathbb{R}$ that is a continuity point of F; the same definition applies in the case when F_n and F are d-dimensional joint distribution functions. Similarly, we say that a sequence $(X_n)_{n=1}^\infty$ of r.v.s (or, more generally, d-dimensional random vectors) converges in distribution to F (a one-dimensional, or more generally d-dimensional, distribution function), and denote $X_n \xrightarrow[n\to\infty]{d} F$, if F_{X_n} converges in distribution to F, where for each n, F_{X_n} denotes the distribution function of X_n.

We will repeatedly encounter a few of the special distributions of probability theory, namely the **geometric**, **exponential** and **Poisson** distributions. The ubiquitous **Gaussian** (a.k.a. **normal**) distribution will also make a couple of brief appearances. For easy reference, here are their definitions.

The geometric distribution. If $0 < p < 1$, we say that an r.v. X has the geometric distribution with parameter p, and denote $X \sim \text{Geom}(p)$, if

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \qquad (k = 1, 2, \ldots).$$

The exponential distribution. If $\alpha > 0$, we say that an r.v. X has the exponential distribution with parameter α, and denote $X \sim \text{Exp}(\alpha)$, if

$$\mathbb{P}(X \geq t) = e^{-\alpha t}, \qquad (t \geq 0).$$

The Poisson distribution. If $\lambda > 0$, we say that an r.v. X has the Poisson distribution with parameter α, and denote $X \sim \text{Poi}(\lambda)$, if

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad (k = 0, 1, 2, \ldots).$$

The Gaussian distribution. If $\mu \in \mathbb{R}$ and $\sigma > 0$, we say that an r.v. X has the Gaussian distribution with mean μ and variance σ^2, and denote $X \sim N(\mu, \sigma^2)$, if

$$\mathbb{P}(a \le X \le b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-(x-\mu)^2/2\sigma}\,dx, \qquad (a < b).$$

0.2 Little-*o* and big-*O* notation

Throughout the book, we are frequently concerned with asymptotic estimates for various quantities as a parameter (usually, but not always, a discrete parameter n) converges to a limit (usually ∞). We use the standard $o(\cdot)$ ("**little-**o") and $O(\cdot)$ ("**big-**O") notation conventions. In the typical case of a discrete parameter n converging to ∞ these are defined as follows. If a_n and b_n are functions of n, the statement

$$a_n = o(b_n) \quad \text{as } n \to \infty$$

means that $\lim_{n\to\infty} a_n/b_n = 0$. The statement

$$a_n = O(b_n) \quad \text{as } n \to \infty$$

means that there exists a constant $M > 0$ such that $|a_n/b_n| \le M$ for all large enough values of n. Similarly, one can define statements such as "$f(x) = O(g(x))$ as $x \to L$" and "$f(x) = o(g(x))$ as $x \to L$"; we leave this variation to the reader to define precisely. Big-O and little-o notation can also be used more liberally in equations such as

$$a_n = \sqrt{n} + O(1) + O(\log n) + o(c_n) \quad \text{as } n \to \infty,$$

whose precise meaning is "$a_n - \sqrt{n}$ can be represented as a sum of three quantities x_n, y_n and z_n such that $x_n = O(1)$, $y_n = O(\log n)$ and $z_n = o(c_n)$." Usually such statements are derived from an earlier explicit description of the x_n, y_n, and z_n involved in such a representation. Frequently several big-O and little-o expressions can be combined into one, as in the equation

$$O(1) + O(\log n) + o(1/n) = O(\log n) \quad \text{as } n \to \infty.$$

As illustrated previously, asymptotic statements are usually accompanied by a qualifier like "as $n \to \infty$" indicating the parameter and limiting value with respect to which they apply. However, in cases when this specification is clear from the context it may on occasion be omitted.

More information regarding asymptotic estimation methods, along with many examples of the use of little-o and big-O notation, can be found in [49], [93].

0.3 Stirling's approximation

The canonical example of an interesting asymptotic relation is Stirling's approximation for $n!$. In the above notation it is written as

$$n! = (1 + o(1))\sqrt{2\pi n}(n/e)^n \quad \text{as } n \to \infty. \tag{0.1}$$

We make use of (0.1) on a few occasions. In some cases it is sufficient to use the more elementary (nonasymptotic) lower bound

$$n! \ge (n/e)^n \qquad (n \ge 1), \tag{0.2}$$

which is proved by substituting $x = n$ in the trivial inequality $e^x \ge x^n/n!$ valid for all $x \ge 0$. The relation (0.1) is harder (but not especially hard) to prove. A few different proofs can be found in [35, Section 6.3], [40], Sections II.9 and VII.3 of [41], [49, Section 9.6], [106], and p. 312 of this book.

1

Longest increasing subsequences in random permutations

Chapter summary. If σ is a permutation of n numbers, we consider the maximal length $L(\sigma)$ of an increasing subsequence of σ. For a permutation chosen *uniformly at random* from among all permutations of order n, how large can we expect $L(\sigma)$ to be? The goal of this chapter is to answer this question. The solution turns out to be rather complicated and will take us on a journey through a fascinating mathematical landscape of concepts such as **integer partitions**, **Young tableaux**, **hook walks**, **Plancherel measures**, **large deviation principles**, **Hilbert transforms**, and more.

1.1 The Ulam–Hammersley problem

We begin with a question about the asymptotic behavior of a sequence of real numbers. Let S_n denote the group of permutations of order n. If $\sigma \in S_n$ is a permutation, a **subsequence** of σ is a sequence $(\sigma(i_1), \sigma(i_2), \ldots, \sigma(i_k))$, where $1 \le i_1 < i_2 < \ldots < i_k \le n$. The subsequence is called an **increasing subsequence** if $\sigma(i_1) < \sigma(i_2) < \ldots < \sigma(i_k)$, a **decreasing subsequence** if $\sigma(i_1) > \sigma(i_2) > \ldots > \sigma(i_k)$, and a **monotone subsequence** if it is either increasing or decreasing. Define $L(\sigma)$ to be the maximal length of an increasing subsequence of σ. That is,

$$L(\sigma) = \max\Big\{1 \le k \le n : \sigma \text{ has an increasing subsequence of length } k\Big\}.$$

Similarly, define $D(\sigma)$ to be the maximal length of a *decreasing* subsequence of σ, i.e.,

$$D(\sigma) = \max\Big\{1 \le k \le n : \sigma \text{ has a decreasing subsequence of length } k\Big\}.$$

For example, if $\sigma = (3, 1, 6, 7, 2, 5, 4)$, then $L(\sigma) = 3$, since it has (several) increasing subsequences of length 3, but no increasing subsequence of length 4. Similarly, one can verify easily that $D(\sigma) = 3$.

Now define the sequence of numbers

$$\ell_n = \frac{1}{n!} \sum_{\sigma \in S_n} L(\sigma), \qquad (n = 1, 2, \ldots).$$

That is, ℓ_n is the average of $L(\sigma)$ over all permutations of order n. For example, the first few values in the sequence are $\ell_1 = 1$, $\ell_2 = 3/2$, $\ell_3 = 2$, $\ell_4 = 29/12$, $\ell_5 = 67/24$. We are interested in the problem of determining the asymptotic behavior of ℓ_n as n grows large. A version of the problem was first mentioned in a 1961 paper by Stanisław Ulam [138], a Polish-American mathematician better known for his work on the hydrogen bomb. In his paper, which concerned the Monte Carlo method for numerical computation (which Ulam pioneered), he discussed briefly the idea of studying the statistical distribution of the maximal monotone subsequence length in a random permutation; this was brought up as an example of the kinds of problems that can be attacked using Monte Carlo calculations. Subsequently, the question came to be referred to as “Ulam’s problem” by some authors—starting with John M. Hammersley, who undertook (with some success) the first serious study of the problem, which he presented in a 1970 lecture and accompanying article [54].[1] To honor Hammersley’s contribution to analyzing and popularizing Ulam’s question, we refer to the problem here as the **Ulam–Hammersley problem**.

In this chapter and the next one we describe the developments leading up to a rather complete solution of Ulam and Hammersley’s problem. The techniques developed along the way to finding this solution did much more than solve the original problem; in fact, they paved the way to many other interesting developments, some of which are described later in the book.

To avoid unnecessary suspense, one form of the “final answer,” obtained in 1998 by Jinho Baik, Percy A. Deift, and Kurt Johansson [11], is as follows: as $n \to \infty$, we have

$$\ell_n = 2\sqrt{n} + cn^{1/6} + o(n^{1/6}), \tag{1.1}$$

where $c = -1.77108\ldots$ is a constant having a complicated definition in terms of the solution to a certain differential equation, the Painlevé equation of type II. We shall have to wait until Chapter 2 to see where this more exotic part of the asymptotics comes from. In this chapter our goal is to prove a first major result in this direction, which identifies only the leading asymptotic term $2\sqrt{n}$. The result, proved by Anatoly Vershik and Sergei Kerov [142], [143] and independently by Benjamin F. Logan and Lawrence A. Shepp [79] in 1977,[2] is the following.

Theorem 1.1 (The asymptotics of ℓ_n) *We have the limit*

$$\frac{\ell_n}{\sqrt{n}} \to 2$$

as $n \to \infty$. Furthermore, the limit is the same for the "typical" permutation of order n. That is, if for each n, σ_n denotes a uniformly random permutation in S_n, then $L(\sigma_n)/\sqrt{n} \to 2$ in probability as $n \to \infty$.

1.2 The Erdős–Szekeres theorem

To gain an initial understanding of the problem, let us turn to a classical result in combinatorics dating from 1935, the Erdős–Szekeres theorem.[3] Paul Erdős and George Szekeres observed that if a permutation has no long increasing subsequence, its elements must in some sense be arranged in a somewhat decreasing fashion, so it must have a commensurately long *decreasing* subsequence. The precise result is as follows.

Theorem 1.2 (Erdős–Szekeres theorem) *If $\sigma \in S_n$ and $n > rs$ for some integers $r, s \in \mathbb{N}$, then either $L(\sigma) > r$ or $D(\sigma) > s$.*

Proof We introduce the following variation on the permutation statistics $L(\cdot)$ and $D(\cdot)$: for each $1 \le k \le n$, let $L_k(\sigma)$ denote the maximal length of an increasing subsequence of σ that ends with $\sigma(k)$, and similarly let $D_k(\sigma)$ denote the maximal length of a decreasing subsequence of σ that ends with $\sigma(k)$.

Now consider the n pairs $(D_k(\sigma), L_k(\sigma))$, $1 \le k \le n$. The key observation is that they are all distinct. Indeed, for any $1 \le j < k \le n$, if $\sigma(j) < \sigma(k)$ then $L_j(\sigma) < L_k(\sigma)$, since we can take an increasing subsequence of σ that ends with $\sigma(j)$ and has length $L_j(\sigma)$, and append $\sigma(k)$ to it. If, on the other hand, $\sigma(j) > \sigma(k)$, then similarly we get that $D_j(\sigma) < D_k(\sigma)$, since any decreasing subsequence that ends with $\sigma(j)$ can be made longer by appending $\sigma(j)$ to it.

The conclusion from this observation is that for some $1 \leq k \leq n$, either $L_k(\sigma) > r$ or $D_k(\sigma) > s$, since otherwise the n distinct pairs $(D_k(\sigma), L_k(\sigma))$ would all be in the set $\{1, 2, \dots, r\} \times \{1, 2, \dots, s\}$, in contradiction to the assumption that $n > rs$. This proves the theorem. □

It is also interesting to note that the condition $n > rs$ in the theorem cannot be weakened. Indeed, it is easy to construct a permutation σ of order exactly rs for which $L(\sigma) = r$ and $D(\sigma) = s$; for example, define $\sigma(si + j) = si - j + s + 1$ for $0 \leq i < r$, $1 \leq j \leq s$ (this permutation has r "blocks," each comprising a decreasing s-tuple of numbers, with the ranges of successive blocks being increasing). In fact, it turns out that the set of permutations that demonstrate the sharpness of the condition has a very interesting structure; this topic is explored further in Chapter 3.

1.3 First bounds

From here on and throughout this chapter, σ_n denotes a uniformly random permutation of order n, so that in probabilistic notation we can write $\ell_n = \mathbb{E}L(\sigma_n)$. We can now use Theorem 1.2 to obtain a lower bound for ℓ_n.

Lemma 1.3 *For all $n \geq 1$ we have*

$$\ell_n \geq \sqrt{n}. \tag{1.2}$$

Proof Rephrasing Theorem 1.2 slightly, we can say that for each permutation $\sigma \in S_n$ we have $L(\sigma)D(\sigma) \geq n$. Now, ℓ_n is defined as the average value of $L(\sigma)$ over all $\sigma \in S_n$. However, by symmetry, clearly it is also the average value of $D(\sigma)$. By linearity of expectations of random variables, this is also equal to

$$\ell_n = \frac{1}{n!} \sum_{\sigma \in S_n} \frac{L(\sigma) + D(\sigma)}{2} = \mathbb{E}\left(\frac{L(\sigma_n) + D(\sigma_n)}{2}\right).$$

By the inequality of the arithmetic and geometric means, we get that

$$\ell_n \geq \mathbb{E}\left(\sqrt{L(\sigma_n)D(\sigma_n)}\right) \geq \sqrt{n}. \qquad \square$$

Comparing (1.2) with (1.1), we see that the bound gives the correct order of magnitude, namely $\sqrt{n}$, for ℓ_n, but with a wrong constant. What about an upper bound? As the following lemma shows, we can also fairly easily get an upper bound of a constant times $\sqrt{n}$, and thus establish that $\sqrt{n}$ is the correct order of magnitude for ℓ_n. This will give us a coarse, but still interesting, understanding of ℓ_n.

Lemma 1.4 *As $n \to \infty$ we have*

$$\limsup_{n\to\infty} \frac{\ell_n}{\sqrt{n}} \le e. \tag{1.3}$$

Proof For each $1 \le k \le n$, let $X_{n,k}$ denote the number of increasing subsequences of the random permutation σ_n that have length k. Now compute the expected value of $X_{n,k}$, noting that this is equal to the sum, over all $\binom{n}{k}$ subsequences of length k, of the probability for that subsequence to be increasing, which is $1/k!$. This gives

$$\mathbb{E}(X_{n,k}) = \frac{1}{k!}\binom{n}{k}.$$

This can be used to bound the probability that $L(\sigma_n)$ is at least k, by noting (using (0.2)) that

$$\begin{aligned} \mathbb{P}(L(\sigma_n) \ge k) = \mathbb{P}(X_{n,k} \ge 1) \le \mathbb{E}(X_{n,k}) &= \frac{1}{k!}\binom{n}{k} \\ &= \frac{n(n-1)\ldots(n-k+1)}{(k!)^2} \le \frac{n^k}{(k/e)^{2k}}. \end{aligned} \tag{1.4}$$

Fixing some $\delta > 0$ and taking $k = \lceil (1+\delta)e\sqrt{n}\rceil$, we therefore get that

$$\mathbb{P}(L(\sigma_n) \ge k) \le \frac{n^k}{(k/e)^{2k}} \le \left(\frac{1}{1+\delta}\right)^{2k} \le \left(\frac{1}{1+\delta}\right)^{2(1+\delta)e\sqrt{n}},$$

a bound that converges to 0 at a rate exponential in $\sqrt{n}$ as $n \to \infty$. It follows (noting the fact that $L(\sigma) \le n$ for all $\sigma \in S_n$) that

$$\begin{aligned} \ell_n = \mathbb{E}(L(\sigma_n)) &\le \mathbb{P}(L(\sigma_n) < k)(1+\delta)e\sqrt{n} + \mathbb{P}(L(\sigma_n) \ge k)n \\ &\le (1+\delta)e\sqrt{n} + O(e^{-c\sqrt{n}}), \end{aligned}$$

where c is some positive constant that depends on δ. This proves the claim, since δ was an arbitrary positive number. □

Note that the proof of Lemma 1.4 actually gave slightly more information than what was claimed, establishing the quantity $(1+\delta)e\sqrt{n}$ as a bound not just for the *average* value of $L(\sigma_n)$, but also for the *typical* value, namely the value that is attained with a probability close to 1 for large n. Furthermore, the bounds we derived also yielded the fact (which will be useful later on) that the probability of large fluctuations of $L(\sigma_n)$ from its typical value decays like an exponential function of $\sqrt{n}$. We record these observations in the following lemma.

Lemma 1.5 *For any $\alpha > e$ we have for all n that*

$$\mathbb{P}(L(\sigma_n) > \alpha\sqrt{n}) \leq Ce^{-c\sqrt{n}}$$

for some constants $C, c > 0$ that depend on α but not on n.

It is interesting to compare this with the argument that was used to prove Lemma 1.3, which really only bounds the average value of $L(\sigma_n)$ and not the typical value, since it does not rule out a situation in which (for example) approximately half of all permutations might have a value of $L(\sigma)$ close to 0 and the other half have a value close to $2\sqrt{n}$. However, as we shall see in the next section, in fact the behavior of $L(\sigma_n)$ for a typical permutation σ_n is asymptotically the same as that of its average value.

1.4 Hammersley's theorem

Our goal in this section is to prove the following result, originally due to Hammersley [54].

Theorem 1.6 (Hammersley's convergence theorem for the maximal increasing subsequence length) *The limit $\Lambda = \lim_{n\to\infty} \frac{\ell_n}{\sqrt{n}}$ exists. Furthermore, we have the convergence $L(\sigma_n)/\sqrt{n} \to \Lambda$ in probability as $n \to \infty$.*

Hammersley's idea was to reformulate the problem of studying longest increasing subsequences in permutations in a more geometric way. Denote by $\preceq$ a partial order on $\mathbb{R}^2$ where the relation $(x_1, y_1) \preceq (x_2, y_2)$ holds precisely if $x_1 \leq x_2$ and $y_1 \leq y_2$. For a set $A = ((x_k, y_k))_{k=1}^n$ of n points in the plane, an **increasing subset** of A is a subset any two of whose elements are comparable in the order $\preceq$ (in the context of partially ordered sets such a subset of A would be called a **chain**). See Fig. 1.1. Denote by $L(A)$ the maximal length of an increasing subset of A. Note that this generalizes the definition of $L(\sigma)$ for a permutation $\sigma \in S_n$, since in that case $L(\sigma) = L(G_\sigma)$, where $G_\sigma = \{(i, \sigma(i)) : 1 \leq i \leq n\}$ is the graph of σ.