

A probabilistic approach to the asymptotics of integer partitions

Richard Hawes

University of Leeds

1 Introduction

A *partition* is a decomposition of a natural number into a sum of unordered summands. More specifically, a sequence $\lambda = (\lambda_1, \lambda_2, \dots)$, with $\lambda_i \in \mathbb{N} = \{1, 2, \dots\}$, $\lambda_1 \leq \lambda_2 \leq \dots$, is called a partition of a natural number n (notation: $\lambda \vdash n$) if

$$n = \lambda_1 + \lambda_2 + \dots.$$

The total number of partitions of n is denoted by $p(n)$ and is called the *partition function*.

A convenient way to represent partitions is to use *multiplicities*. The multiplicity of a number is how many times it occurs in a partition. Let $r_x(\lambda)$ denote the multiplicity of x in the partition λ . Multiplicity has the following simple properties:

$$\sum_{x=1}^{\infty} r_x(\lambda) = \#\{\lambda_i \in \lambda\}, \quad \sum_{x=1}^{\infty} x r_x(\lambda) = n. \quad (1.1)$$

For example, in the partition $12 = 1 + 1 + 2 + 2 + 2 + 4$, we have $r_1(12) = 2$, $r_2(12) = 3$, $r_4(12) = 1$ and $r_i(12) = 0 \forall i \neq \{1, 2, 4\}$.

Partitions were first studied by Leonhard Euler. He found a formula for the generating function of the number of partitions (see Andrews, 1976, for more details on using generating functions for partitions)

$$F(s) := 1 + \sum_{n=1}^{\infty} p(n)s^n = \prod_{x=1}^{\infty} \frac{1}{1-s^x}.$$

To see this, first notice that each fraction is the sum of an infinite geometric progression, so that

$$\frac{1}{1-s} \cdot \frac{1}{1-s^2} \cdot \frac{1}{1-s^3} \cdots = (1+s+s^2+\cdots)(1+s^2+s^4+\cdots)(1+s^3+s^6+\cdots)\cdots$$

By expanding these parentheses we find s^n can be written in the form $s^{r_1+2r_2+3r_3+\cdots}$, which implies that $n = r_1 + 2r_2 + 3r_3 + \cdots$. Thus the second formula of (1) is satisfied, which means that each r_i is the multiplicity of i in a partition of n . Therefore each solution of this equation represents a unique partition of n . Hence the coefficient of s^n is the total number of partitions of n .

Further progress in the field of partitions was due to application of powerful analytical methods by Hardy, Ramanujan and Rademacher. Early in the twentieth century Hardy and Ramanujan studied the following equation

$$p(n) = \frac{1}{2\pi i} \oint_{\gamma} \frac{F(s)}{s^{n+1}} ds,$$

where γ is, say, a circle centred at the origin that contains the unit circle. This integral is very difficult to evaluate due to the many poles created by the generating function. Hardy and Ramanujan managed to find many results about the partition function but they did not manage to obtain an exact solution. Their result for $p(n)$ from this study produced a value of $p(n)$ with an absolute error of less than a half. Rademacher finished the work (see Andrews, 1976) by finding a fast convergent series that gave exact values for $p(n)$. The first term of the series gives the following asymptotic formula:

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp \left\{ \frac{2\pi}{\sqrt{6}} \sqrt{n} \right\} \quad (n \rightarrow \infty).$$

2 Probabilistic approach

Producing the first term of the series using complex integration is quite problematic. We regain this result for $p(n)$ in a much simpler and efficient way by applying some probability to the situation. Denote

$$\mathcal{L}_n := \{\lambda : \lambda \vdash n\}, \quad \mathcal{L} := \bigcup_n \mathcal{L}_n.$$

Thus \mathcal{L}_n consists of all partitions of n and \mathcal{L} is the whole space of partitions. We now introduce probability by applying the uniform measure to \mathcal{L}_n . This is the most natural choice of measure since partitions are unique combinatorial objects and we therefore have no reason to favour one over another. Let

$$\mathbb{P}_n(\lambda) := \frac{1}{p(n)}, \quad \lambda \in \mathcal{L}_n. \quad (2.2)$$

Now we have a formula which involves $p(n)$ but no way to calculate it. The probabilistic approach suggested by Vershik (1996) is to represent the probability distribution of \mathbb{P}_n on \mathcal{L}_n as a conditional distribution induced by some probability measure \mathbb{Q} defined on the whole space \mathcal{L} . More specifically, we want to construct the measure \mathbb{Q} so that

$$\mathbb{Q}(\lambda \mid \lambda \in \mathcal{L}_n) \equiv \frac{\mathbb{Q}(\lambda)}{\mathbb{Q}(\mathcal{L}_n)} = \mathbb{P}_n(\lambda), \quad \lambda \in \mathcal{L}_n. \quad (2.3)$$

Constructing the measure so (3) is true we have by comparing (2) and (3) that

$$p(n) = \frac{\mathbb{Q}(\mathcal{L}_n)}{\mathbb{Q}(\lambda)}, \quad \lambda \in \mathcal{L}_n. \quad (2.4)$$

The procedure from here is to construct \mathbb{Q} to satisfy (3) and then gain approximations for the two probabilities in (4). This will result in an asymptotic formula for $p(n)$.

3 Construction of the measure \mathbb{Q}

Consider a sequence $R = (R_x) = (R_1, R_2, \dots)$ of independent random variables with distributions

$$\mathbb{Q}\{R_x = k\} = z^{xk}(1 - z^x), \quad k = 0, 1, 2, \dots,$$

where z is a parameter, $0 < z < 1$. Note that these are geometric distributions with probability of success $1 - z^x$ and probability of failure z^x . We want this sequence to represent a partition. By letting R_x represent the multiplicity of x we form a partition of $\sum_x xR_x$ when we sum the R_x . We interpret $R \in \mathcal{L}_n$ to mean that $\sum_x xR_x = n$. This choice of distribution induces

uniformity. Also the random variables in the sequence R are finitely supported which is a necessary condition since this guarantees that when summed the sequence of multiplicities R will always form a partition of a natural number.

Lemma 3.1. $\mathbb{Q}\{R_x > 0 \text{ finitely often}\} = 1$

Proof. Apply the Borel-Cantelli lemma:

$$\sum_x \mathbb{Q}\{R_x > 0\} = \sum_x z^x = \frac{z}{1-z} < \infty.$$

□

Lemma 3.2. *The conditional measure $\mathbb{Q}\{\cdot \mid R \in \mathcal{L}_n\}$ is uniform on \mathcal{L}_n .*

Proof. We have,

$$\begin{aligned} \mathbb{Q}\{R = (r_1, r_2, \dots) \mid R \in \mathcal{L}_n\} &= \frac{\mathbb{Q}(R = (r_1, r_2, \dots))}{\mathbb{Q}(\mathcal{L}_n)} = \frac{\prod_x \mathbb{Q}(R_x = r_x)}{\sum_{R \in \mathcal{L}_n} \prod_x \mathbb{Q}(R_x = r_x)} \\ &= \frac{\prod_x z^{x r_x} (1 - z^x)}{\sum_{R \in \mathcal{L}_n} \prod_x z^{x r_x} (1 - z^x)} = \frac{z^{\sum_x x r_x}}{\sum_{R \in \mathcal{L}_n} z^{\sum_x x r_x}} \\ &= \frac{z^n}{\sum_{R \in \mathcal{L}_n} z^n} = \frac{1}{\#(\mathcal{L}_n)} = \frac{1}{p(n)}. \end{aligned}$$

□

Thus there is no dependence on the parameter z .

4 Choice of z

To choose z we impose the following condition upon expectation (motivated by how we chose our partition from the sequence R):

$$\mathbb{E} \sum_x x R_x = \sum_x \frac{x z^x}{1 - z^x} \sim n \quad (n \rightarrow \infty).$$

Set

$$\alpha := \frac{\delta}{\sqrt{n}} \rightarrow 0 \quad (n \rightarrow \infty)$$

and seek z in the form

$$z := e^{-\alpha}.$$

Then

$$\sum_x \frac{x z^x}{1 - z^x} = \frac{1}{\alpha^2} \sum_x \alpha f(\alpha x) \sim \int_0^\infty f(u) du \quad (\alpha \rightarrow 0),$$

where

$$f(u) = \frac{u e^{-u}}{1 - e^{-u}}.$$

This integral can be evaluated and we find that

$$\mathbb{E} \sum_x x R_x = \frac{\pi^2}{6\alpha^2} \implies \delta = \frac{\pi}{\sqrt{6}}.$$

5 Solution of $p(n)$

By (4) and the proof of Lemma 2 we have

$$p(n) = \frac{\mathbb{Q}(\mathcal{L}_n)}{\mathbb{Q}(R = (r_1, r_2, \dots))} = \frac{\mathbb{Q}\{\sum_x xR_x = n\}}{z^n \prod_x (1 - z^x)}. \quad (5.5)$$

Now

$$z^n \prod_x (1 - z^x) = \exp\left(-n\alpha + \sum_x \ln(1 - e^{-\alpha x})\right).$$

Using the Euler-Macluarin summation formula (see De Bruijn, 1970), it can be shown that

$$\sum_x \ln(1 - e^{-\alpha x}) = -\frac{\pi^2}{6\alpha} - \log \sqrt{\alpha} + \log \sqrt{2\pi} + O(\alpha),$$

and it follows that

$$\mathbb{Q}(R = (r_1, r_2, \dots)) \sim \sqrt{2} 6^{1/4} n^{1/4} \cdot \exp\left\{\frac{-2\pi}{\sqrt{6}} \sqrt{n}\right\} \quad (n \rightarrow \infty) \quad (5.6)$$

The numerator in (5) is evaluated via an appropriate local central limit theorem, which can be proved using the method of characteristic functions. The local central limit theorem yields (cf. Freiman *et al.*, 1998)

$$\mathbb{Q}\left\{\sum_x xR_x = n\right\} \sim \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{(n - a_n)^2}{2\sigma_n^2}\right), \quad (5.7)$$

where

$$a_n := \mathbb{E} \sum_x xR_x, \quad \sigma_n^2 := \text{Var} \sum_x xR_x.$$

The variance σ_n^2 is evaluated in a similar fashion to a_n :

$$\sigma_n^2 = \sum_x \frac{x^2 e^{-\alpha x}}{(1 - e^{-\alpha x})^2} \sim \frac{2\sqrt{6}}{\pi} n^{3/2}.$$

Substituting the values of a_n and σ_n^2 into (7) it follows that

$$\mathbb{Q}\left\{\sum_x xR_x = n\right\} \sim \frac{n^{-3/4}}{2 \cdot 6^{1/4}}. \quad (5.8)$$

Substituting (6) and (8) into (5) achieves Hardy and Ramanujan's asymptotic formula for $p(n)$.

References

- Andrews, G. E. (1976). *The Theory of Partitions*. (Encyclopedia of Mathematics and its Applications, Vol. 2). Addison Wesley, Reading, MA.
- De Bruijn, N. G. (1970). *Asymptotic Methods in Analysis*, 3rd ed., (Bibliotheca Mathematica). North-Holland, Amsterdam.
- Freiman, G., Vershik, A. M. and Yakubovich, Yu. M. (1998). A local limit theorem for random strict partitions, *Theory Probab. Appl.*, **44**, 453-468.
- Vershik, A. M. (1996). Statistical mechanics of combinatorial partitions, and their limit configurations. *Funct. Anal. Appl.*, **30**, 90-105.