

On the number of summands in a random prime partition

Dimbinaina Ralaivaosaona

Abstract. We study the length (number of summands) in partitions of an integer into primes, both in the restricted (unequal summands) and unrestricted case. It is shown how one can obtain asymptotic expansions for the mean and variance (and potentially higher moments), which is in contrast to the fact that there is no asymptotic formula for the number of such partitions in terms of elementary functions. Building on ideas of Hwang, we also prove a central limit theorem in the restricted case. The technique also generalizes to partitions into powers of primes, or even more generally, the values of a polynomial at the prime numbers.

Mathematics Subject Classification (2010). Primary 05A17; Secondary 11P82.

Keywords. Asymptotic expansions, limit distribution, Mellin transform, prime partitions, saddle point method.

1. Introduction

Prime partitions (partitions into primes) of an integer n are ways to write n as a sum of primes. We say that a partition is restricted if repetitions are not allowed, and unrestricted otherwise.

For ordinary partitions (partitions into arbitrary summands), Erdős and Lehner [4] proved that the limit distribution for the number of summands is Gaussian in the restricted case and the so-called *double exponential* (Gumbel) distribution in the unrestricted case. In their proof they made use of the well known Hardy-Ramanujan asymptotic formula [9] for the number of integer partitions:

$$P(n) \sim \frac{1}{4n\sqrt{3}} \exp(\pi\sqrt{2n/3})$$

This project is supported by the German Academic Exchange Service (DAAD), in association with the African Institute for Mathematical Sciences (AIMS). Code No.:A/09/04406.

as $n \rightarrow \infty$. The analogue of this result for primes is quite complicated as the asymptotic formula cannot be expressed in terms of elementary functions, see [18] for more details. However, in this case Hardy and Ramanujan [8, 9] proved the following asymptotic formula

$$\log p(n) \sim 2\pi\sqrt{n/(3\log n)},$$

where $p(n)$ is the number of partitions of n into primes.

There are also more general results concerning the number of summands in so-called λ -partitions (unrestricted), which are partitions whose summands belong to a given increasing sequence of positive integers λ . The sequence λ must satisfy some technical conditions, and it has been shown that one can apply the result to the sequence of primes (see Section 4). This was found in a paper by Haselgrove and Temperley in [10] and later improved independently by Lee [13] and Richmond [15]. Their theorems give a generalization of the Erdős-Lehner result for unrestricted partitions.

On the other hand, Hwang developed general theorems for restricted λ -partitions, showing that the limit distribution remains Gaussian under certain technical conditions (see [12] for details). Efforts have been made to reduce these conditions, see for instance [14], but yet the sequence of primes fails to satisfy these conditions, as also mentioned explicitly in Hwang's paper.

With some modifications of the method used in [12] one can prove the following for primes:

Theorem 1. *The number of summands in a random restricted partition of n into distinct primes is asymptotically normally distributed with mean and variance satisfying the following asymptotic formulas:*

$$\mu_n = \frac{2\log 2}{\pi} \sqrt{\frac{6n}{\log n}} \left(1 - \frac{\log \log n}{2\log n} + O\left(\frac{1}{\log n}\right)\right)$$

and

$$\sigma_n^2 = \frac{\sqrt{6}}{\pi} \left(1 - \frac{12\log^2 2}{\pi^2}\right) \sqrt{\frac{n}{\log n}} \left(1 - \frac{\log \log n}{2\log n} + O\left(\frac{1}{\log n}\right)\right)$$

respectively as $n \rightarrow \infty$.

Further examples of central limit theorems in the context of partitions include those by Goh and Schmutz [7] for the number of distinct parts and by Brennan, Knopfmacher and Wagner [2] for ascents of size d or more (equivalently, parts of multiplicity d or more).

In principle, the presented method can be used to determine even more terms of an asymptotic expansion. Let us mention that the main asymptotic term of μ_n (and also of mean and variance in the unrestricted case) already occurs in [16], but with the factor $\log n$ missing.

We shall prove this theorem in detail in Section 3, and then later on, in Section 4 we give an asymptotic formula for mean and variance for the distribution of the number of summands in a random unrestricted prime partition of n . Finally, in Section 5, we discuss how our results can be generalized to a wider variety of sequences involving primes.

2. Definitions and preliminary results

In this section we gather important information about primes that are fundamental for the proof of our main results. But first, let us agree on notations and conventions that will be used throughout the paper.

Notation. Assume that all partitions of n into distinct primes are equally likely. Let ϖ_n denote the number of summands in a random partition, the mean and standard deviation are denoted by μ_n and σ_n respectively. The random variable ϖ_n^* , its mean μ_n^* and variance σ_n^* are defined analogously for unrestricted prime partitions. We shall use \prod_p and \sum_p as abbreviations for the product and sum over all primes respectively. The Dirichlet series associated to the sequence of primes is defined by

$$D(s) := \sum_p p^{-s}$$

for complex numbers s with $\operatorname{Re}(s) > 1$.

The first result that we need is the following about the exponential sum

$$g(\tau) := \sum_p e^{-p\tau}.$$

The result states that:

Lemma 2. *For any constant $1/3 < c < 1/2$ there is a constant $c_1 > 0$ such that*

$$g(r) - \operatorname{Re} g(\tau) \geq c_1 \log^2 \frac{1}{r}$$

for $\tau = r + iy$ with $r^{1+c} \leq |y| \leq \pi$ as $r \rightarrow 0^+$.

A major part of this result has already been proved by Roth and Szekeres in [17, last section]. Since it is a fundamental result and its proof is not too long, we shall give a complete proof here.

Proof. First, let us assume that $\pi r \leq |y| \leq \pi$. We have

$$g(r) - \operatorname{Re} g(\tau) = \sum_p e^{-pr} (1 - \cos py) \tag{1}$$

$$\geq \sum_{p \leq r^{-1}} e^{-pr} (1 - \cos p|y|) \tag{2}$$

$$\geq 8e^{-1} \sum_{p \leq r^{-1}} \left\| \frac{p|y|}{2\pi} \right\|^2 \tag{3}$$

where $\|\cdot\|$ denotes the distance from the nearest integer. To simplify our notation let us define $\alpha = \alpha(y)$ to be $|y|/2\pi$, then we have $r/2 \leq |\alpha| \leq 1/2$. First, if α is rational, say $\alpha = a/q$, where a and q (with $q > 1$) are coprime, then

$$\sum_{p \leq r^{-1}} \left\| \frac{pa}{q} \right\|^2 \gg \sum_{p \leq r^{-1}} \frac{1}{q^2} \gg \frac{1}{q^2 r \log 1/r}$$

for sufficiently small r . This is sufficient for small values of q , say $q \leq r^{-1/3}$. Now suppose that $2r^{-1} \geq q \geq r^{-1/3}$. Then there are $\gg 1/(r \log 1/r)$ elements in $\{pa : p \leq r^{-1}\}$, and each residue class modulo q contains at most $\lceil \frac{1}{qr} \rceil$ elements from this set. It follows that at least $c_2 q / (\log(1/r))$ distinct residue classes occur in $\{pa : p \leq r^{-1}\}$, where c_2 is a positive constant. Therefore, the sum in (3) is at least

$$\frac{1}{q^2} \sum_{j=1}^{\lfloor c_3 q / \log 1/r \rfloor} j^2 > c_4 \frac{r^{-1/3}}{\log^3 1/r}$$

for some positive constants c_3 and c_4 . That settles the case where α is a rational of the form a/q with $q \leq 2r^{-1}$. Otherwise we approximate α by rational in the following way: we choose relatively prime integers a and q such that $q \leq 2\lfloor r^{-1} \rfloor$ and

$$\left| \alpha - \frac{a}{q} \right| \leq \frac{1}{2q\lfloor r^{-1} \rfloor}.$$

Then we claim that for any $p \leq r^{-1}$

$$\|p\alpha\| \geq \frac{1}{2} \|pa/q\|.$$

The claim follows from the fact that if

$$|x - y| \leq \frac{\|y\|}{2}$$

then $\|x\| \geq \frac{\|y\|}{2}$, using the triangle inequality for $\|\cdot\|$. Now the desired estimate follows from the case of rational α .

For the remaining part, that is for $r^{1+c} \leq |y| \leq \pi r$

$$g(r) - \operatorname{Re} g(\tau) = \sum_p e^{-pr} (1 - \cos py) \geq \sum_{p \leq r^{-1}} e^{-pr} (1 - \cos p|y|).$$

The latter sum can be estimated as follows: since $p \leq r^{-1}$ we have $p|y| \leq \pi$, and so

$$1 - \cos p|y| \geq \frac{2}{\pi^2} p^2 y^2.$$

Therefore,

$$\sum_{p \leq r^{-1}} e^{-pr} (1 - \cos p|y|) \geq \sum_{p \leq r^{-1}} e^{-1} \frac{2}{\pi^2} p^2 y^2 \geq c_1 r^{-1+2c} / \log \frac{1}{r},$$

which completes the proof. \square

One of the methods that we are using here is the so-called Mellin transform method which is quite powerful in determining asymptotic formulas for harmonic sums. Our main interest here are harmonic series over primes for which the Dirichlet series $D(s)$ plays a fundamental role. First of all, the

series $D(s)$ is absolutely convergent in the half-plane $\operatorname{Re}(s) > 1$ and therefore analytic in that region. But we can also express $D(s)$ as

$$D(s) = \log \zeta(s) + \sum_p (\log(1 - p^{-s}) + p^{-s}) \quad (4)$$

where $\zeta(s)$ is the Riemann zeta function, so $D(s)$ can be continued analytically to some bigger cut-plane not containing any zeros of $\zeta(s)$. The sum on the right-hand side of (4) is absolutely convergent for $\operatorname{Re}(s) > 1/2$.

Note that if $\hat{f}(s)$ is the Mellin transform of a function $f(x)$ then the Mellin inversion formula says that

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \hat{f}(s)x^{-s} ds \quad (5)$$

for any real c in the fundamental strip of $\hat{f}(s)$. The next lemma gives us asymptotic formulas for some integrals of type (5).

Lemma 3. *Let $F(s)$ be an analytic function in $\operatorname{Re}(s) > 1/2$ admitting the following Taylor expansion around $s = 1$:*

$$F(s) = a_0 + \sum_{k \geq 1}^N \frac{a_k}{k!} (s-1)^k + O((s-1)^{N+1}),$$

and assume furthermore that $F(\sigma + it)$ decays exponentially when $|t| \rightarrow \infty$, uniformly for $\delta^{-1} \leq \sigma \leq \delta$ for some fixed $\delta > 1$. Then we have

$$\frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} F(s)D(s)r^{-s} ds = \sum_{k=0}^N \frac{(-1)^k a_k}{r(\log \frac{1}{r})^{k+1}} + O(\log \log \frac{1}{r} / (r(\log \frac{1}{r})^{N+2}))$$

for any $c > 1$.

If $F(s)$ is meromorphic admitting only a single pole at $s = 1$ with residue 1 and if the rest of the above conditions are satisfied then we have

$$\frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} F(s)D(s)r^{-s} ds = \frac{\log \log \frac{1}{r}}{r} + \frac{B_1}{r} + O(1/(r \log \frac{1}{r}))$$

where B_1 is Mertens's constant, defined as

$$B_1 := \gamma + \sum_p (\log(1 - 1/p) + 1/p),$$

where γ is the Euler–Mascheroni constant.

Similar results are quite common in this context, so we do not give detailed proof here and only provide the main steps.

Proof. The Dirichlet series $D(s)$ admits the following bound

$$|D(\sigma + it)| = O(\log \log(t))$$

for large t , uniformly for $\sigma \geq 1$. This comes from the fact that $\zeta(\sigma + it)$ is bounded above and below by powers of $\log t$ for $\sigma \geq 1$. Therefore, the function $|F(s)D(s)|$ decays exponentially along a vertical line and uniformly for $\operatorname{Re}(s) \geq 1$, so we may safely shift the path of integration to the left

without changing the value of the integral. Hence, for sufficiently small r we may consider the path P defined as the union of the following parts (see Figure 1):

- $P_0 = \{s = 1 + \frac{1}{\log \frac{1}{r}} e^{i\theta} : -\pi/2 \leq \theta \leq \pi/2\} \cup \{s = 1 + x \pm i \frac{1}{\log \frac{1}{r}} : \frac{-1}{\sqrt{\log \frac{1}{r}}} \leq x \leq 0\}$,
- $P_1 = \{s = 1 - \frac{1}{\sqrt{\log \frac{1}{r}}} + it : \frac{1}{\log \frac{1}{r}} \leq |t| \leq \log \frac{1}{r}\}$,
- $P_2 = \{s = 1 + it : |t| \geq \log \frac{1}{r}\} \cup \{s = 1 + x \pm i \log \frac{1}{r} : \frac{-1}{\sqrt{\log \frac{1}{r}}} \leq x \leq 0\}$.

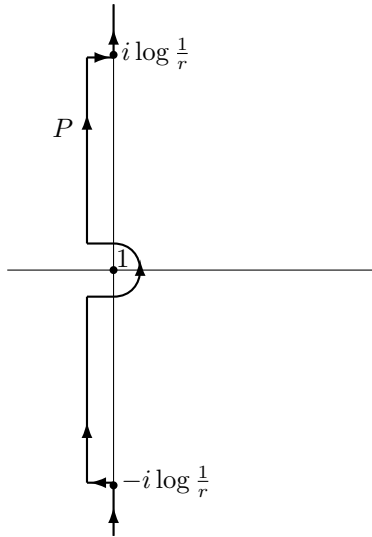


FIGURE 1. Path of integration P .

For r sufficiently small, this path is included in the interior of the zero-free region of the Riemann zeta function (for more details on the zero-free region of the Riemann zeta function see [6]).

On P_1 , one can say that $|D(s)|$ also grows slowly, since the logarithmic derivative $\zeta'(\sigma + it)/\zeta(\sigma + it)$ is bounded above by a power of $\log t$ close to the line $\text{Re}(s) = 1$ (see for instance [1]). Therefore, the contribution from the integrals over P_1 and P_2 are both exponentially small in $\log 1/r$.

On P_0 , one uses Taylor expansion for the first case, or Laurent expansion for the second case of the function $F(s)$ around $s = 1$. Thus the integrals that we need to compute are integrals of the form

$$\frac{1}{2\pi i} \int_{P_0} \log(s-1)(s-1)^k r^{-s} ds$$

for $k = -1, 0, 1, 2, 3, \dots$. Then use the change of variable

$$s = 1 + \frac{z}{\log \frac{1}{r}}.$$

to change the path of integration into the so-called Hankel contour (see for example [5]). The rest of the proof consists of straightforward calculations. \square

3. Proof of the main theorem

To obtain the central limit theorem, we basically follow the ideas in Hwang's paper [12] but first we would like to compute asymptotic formulas for the mean and variance of the random variable ϖ_n . Recall that restricted partitions are partitions without repetitions, so the following bivariate generating function is the generating function for the number of restricted partitions with given length:

$$Q(u, z) = \prod_p (1 + uz^p),$$

i.e., the coefficient of $u^k z^n$ in $Q(u, z)$ is the number of ways of writing n as a sum of exactly k distinct primes. In other words

$$[z^n]Q(u, z) = q(n)\mathbb{E}(u^{\varpi_n})$$

where $q(n) = [z^n]Q(1, z)$ is the total number of ways of writing n as sum of distinct primes. Note that for any u in a fixed bounded interval containing 1, the infinite product $Q(u, z)$ is convergent for $|z| < 1$ and so it is analytic in the unit disc. Let us then define the following function

$$f(u, \tau) := \log Q(u, e^{-\tau}) = \sum_p \log(1 + ue^{-p\tau}), \quad (6)$$

where we always use the principal branch of the logarithm function. The next lemma provides an asymptotic formula for $q(n)$ for large n .

Proposition 4. *The number of unequal partitions of n into primes has the asymptotic formula*

$$q(n) = \frac{e^{nr}}{\sqrt{2\pi f_2(r)}} Q(1, e^{-r})(1 + O(n^{-1/7})) \quad (7)$$

as $n \rightarrow \infty$, where $r > 0$ is the solution of the equation

$$n = \sum_p \frac{pe^{-pr}}{1 + e^{-pr}}. \quad (8)$$

Proof. We know that $q(n)$ is the coefficient of z^n in $Q(1, z)$, so

$$q(n) = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(nit + f(1, r + it)) dt, \quad (9)$$

for any $r > 0$. Then we choose r as defined in Equation (8). The series in (8) is a monotone decreasing function of r , so the solution $r = r(n)$ of (8)

exists and is unique, and it tends to zero as n tends to infinity. Now, we split the integral in (9) as follows: first the integral in the center for $|t| \leq r^{1+\beta}$ and then the tails for $r^{1+\beta} < |t| \leq \pi$, where β is any constant such that $1/3 < \beta < 1/2$. For $|t| \leq r^{1+\beta}$ the function $f(1, r + it)$ admits an expansion

$$f(1, r + it) = f(1, r) + if_1(r)t - f_2(r)\frac{t^2}{2} + O(t^3 \sup_{0 \leq t_0 \leq t} |f_3(r + it_0)|)$$

where $f_k(r)$ denotes the k th derivative of $f(1, \tau)$ with respect to τ at $\tau = r$. We use Lemma 3 to find an asymptotic formula for $f_k(r)$. The Mellin transform of $f_k(r)$ is given by

$$\mathcal{M}(f_k(r), s) = (-1)^{k+1} \text{Li}_{s-k+1}(-1)\Gamma(s)D(s-k),$$

where the function $\text{Li}_s(-1)$ is the polylogarithm function regarded as a function of s . For $\text{Re}(s) > 0$ one can represent $\text{Li}_s(-1)$ as

$$\text{Li}_s(-1) = \sum_{k \geq 1} \frac{(-1)^k}{k^s}$$

and it can be continued analytically to the whole s -plane as

$$\text{Li}_s(-1) = (2^{1-s} - 1)\zeta(s).$$

Therefore by Lemma 3 one gets

$$f_k(r) = (-1)^k k! \frac{\pi^2/12}{r^{k+1} \log \frac{1}{r}} \left(1 + O\left(\frac{1}{\log \frac{1}{r}}\right)\right) \quad (10)$$

as $r \rightarrow 0$. To estimate the third derivative $f_3(r + it)$ we have

$$|f_3(r + it)| \ll \sum_p \frac{p^3 e^{-pr}}{|1 + e^{-p(r+it)}|^3}.$$

For any prime p and $|t| \leq r^{1+\beta}$ we have

$$|1 + e^{-p(r+it)}| \geq 1 - e^{-1}$$

since if $p \geq r^{-1}$, then

$$|1 + e^{-p(r+it)}| \geq 1 - e^{-pr} \geq 1 - e^{-1}$$

and if $p < r^{-1}$ then

$$|1 + e^{-p(r+it)}| \geq 1 + \text{Re}(e^{-p(r+it)}) = 1 + e^{-pr} \cos(pt) > 1.$$

Therefore

$$|f_3(r + it)| \ll \sum_p p^3 e^{-pr} \ll \frac{1}{r^4 \log 1/r}$$

Hence one has

$$nit + f(1, r + it) = f(1, r) - f_2(r)\frac{t^2}{2} + O(r^{3\beta-1}/\log \frac{1}{r}).$$

Thus

$$\int_{-r^{1+\beta}}^{r^{1+\beta}} e^{nit+f(1, r+it)} dt = \int_{-r^{1+\beta}}^{r^{1+\beta}} e^{-f_2(r)t^2/2} dt (1 + O(r^{3\beta-1}/\log \frac{1}{r})),$$

and

$$\int_{-r^{1+\beta}}^{r^{1+\beta}} e^{-f_2(r)t^2/2} dt = \int_{-\infty}^{+\infty} e^{-f_2(r)t^2/2} dt + 2 \int_{r^{1+\beta}}^{+\infty} e^{-f_2(r)t^2/2} dt.$$

The first integral on the right hand side gives the asymptotic formula, the second integral is smaller than any power of r , as $r \rightarrow 0$. It remains to show that the tails are small, and for that we make use of Lemma 2. In fact,

$$\begin{aligned} \frac{|Q(1, e^{-(r+it)})|^2}{Q(1, e^{-r})^2} &= \prod_p \left(1 - \frac{2e^{-pr}(1 - \cos(pt))}{(1 + e^{-pr})^2} \right) \\ &\leq \exp \left(-\frac{1}{2} \sum_p e^{-pr}(1 - \cos(pt)) \right). \end{aligned}$$

Thus, for $r^{1+\beta} < |t| \leq \pi$ Lemma 2 applies and we deduce that the tails of the integral in (9) are exponentially smaller than the main term. Then the asymptotic formula in Equation (7) follows by choosing any $\beta > 3/7$. \square

Mean and variance. The mean μ_n of the random variable ϖ_n can be obtained as

$$\mu_n = \frac{\partial}{\partial u} \mathbb{E}(u^{\varpi_n})|_{u=1}.$$

So we may express μ_n in terms of $Q(u, z)$ as follows:

$$q(n)\mu_n = [z^n]Q(1, z) \sum_p \frac{z^p}{1 + z^p}. \quad (11)$$

By the Cauchy formula we have

$$q(n)\mu_n = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(nit + f(1, r + it))g(r + it)dt \quad (12)$$

for any $r > 0$, where

$$g(\tau) := \sum_p \frac{e^{-p\tau}}{1 + e^{-p\tau}}.$$

We choose $r = r(n)$ as defined in Proposition 4. We shall now estimate the following integral rather than working directly on the integral in (12)

$$q(n)(\mu_n - g(r)) = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(nit + f(1, r + it))(g(r + it) - g(r))dt. \quad (13)$$

In fact, we want to show that the integral in Equation (13) is of small order compared to the order of $g(r)$. We apply the saddle point technique again to the integral (13), it is not hard to show that the tails here are also small. So the main term comes from the integral in the center for which $|t| \leq r^{1+\beta}$ with the same β defined above. Then $g(r + it)$ admits the following expansion

$$g(r + it) - g(r) = ig_1(r)t - g_2(r)\frac{t^2}{2} + O(t^3 \sup_{0 \leq t_1 \leq t} |g_3(r + it_1)|). \quad (14)$$

By means of the Mellin transform method we can show that the k th derivative $g_k(r)$ of $g(\tau)$ satisfies the following asymptotic formula

$$g_k(r) = (-1)^k k! \frac{\log 2}{r^{k+1} \log \frac{1}{r}} \left(1 + O\left(\frac{1}{\log \frac{1}{r}}\right) \right)$$

for small r . One can also prove in a similar way as we did for $f_3(r + it)$, that the error term in Equation (14) is a $O(r^{3\beta-1}/\log \frac{1}{r})$. On the other hand we have

$$e^{nit+f(1,r+it)} = e^{-f_2(r)t^2/2} \left(1 - if_3(r) \frac{t^3}{6} + O(r^{6\beta-2}/\log \frac{1}{r}^2) \right).$$

Therefore, as in the proof of Proposition 4, we can extend the integration to the whole range of real numbers, and we deduce that

$$q(n)(\mu_n - g(r)) = q(n) \left(\frac{f_3(r)g_1(r) - f_2(r)g_2(r)}{2f_2(r)^2} + O(r^{7\beta-3}/\log \frac{1}{r}) \right).$$

To make the error term small, we choose $3/7 < \beta < 1/2$, which implies that

$$\mu_n = \sum_p \frac{e^{-pr}}{1 + e^{-pr}} + \frac{3 \log 2}{\pi^2} + O\left(\frac{1}{\log \frac{1}{r}}\right). \quad (15)$$

Note that we have the following estimate for r :

$$r = r(n) = \frac{\pi}{\sqrt{6n \log n}} \left(1 - \frac{\log \log n}{2 \log n} + O\left(\frac{1}{\log n}\right) \right)$$

as $n \rightarrow \infty$. It is possible to expand the formula even further and get more terms in the expansion. Then the estimate of μ_n in (15) implies the formula in Theorem 1. We do the same for the variance, which is given by the formula

$$\sigma_n^2 = \frac{\partial^2}{\partial^2 u} \mathbb{E}(u^{\varpi_n})|_{u=1} - \mu_n^2 + \mu_n.$$

The second derivative of the function $Q(u, z)$ with respect to u is

$$\frac{\partial^2}{\partial^2 u} Q(u, z)|_{u=1} = Q(1, z) \left(\sum_p \frac{z^p}{1 + z^p} \right)^2 - Q(1, z) \sum_p \frac{z^{2p}}{(1 + z^p)^2}.$$

One can derive the following integral, using equation (11) followed by the Cauchy theorem:

$$q(n)(\sigma_n^2 + \mu_n^2) = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(int + f(1, r + it))(g(r + it)^2 + h(r + it))dt,$$

where

$$h(\tau) = \sum_p \frac{e^{-p\tau}}{(1 + e^{-p\tau})^2}.$$

Then by same method that we used for μ_n , one may show that the variance satisfies the asymptotic formula

$$\sigma_n^2 = h(r) - \frac{g_1(r)^2}{f_2(r)} + O(r^{3\beta-2}/\log^2 \frac{1}{r}).$$

which in turn implies the formula for σ_n^2 in Theorem 1.

Distribution function

Just like the mean and variance, one can also represent the moment generating function of the normalized random variable $(\varpi_n - \mu_n)/\sigma_n$ in terms of $Q(u, z)$. The moment generating function is by definition

$$M_n(t) = \mathbb{E}(e^{(\varpi_n - \mu_n)t/\sigma_n}) \quad (16)$$

$$= \exp\left(-\frac{\mu_n t}{\sigma_n}\right) \frac{Q_n(e^{t/\sigma_n})}{Q_n(1)} \quad (17)$$

where $Q_n(u) = [z^n]Q(u, z)$, and so $Q_n(1) = q(n)$. We shall study the behavior of $Q_n(u)$ for u in a fixed bounded interval containing 1, say $1 - \delta \leq u \leq 1 + \delta$ for a fixed small $\delta > 0$. Throughout this section we will always assume that u is as such, and say that an approximation is uniform in u if it is uniform for u in that interval. We start by an analogue of Proposition 4.

Proposition 5. *The following asymptotic formula holds for the coefficient of z^n in $Q(u, z)$:*

$$Q_n(u) = \frac{1}{\sqrt{2\pi f_2(u, r)}} e^{nr + f(u, r)} \left(1 + O\left(n^{-1/7}\right)\right).$$

uniformly in u , as $n \rightarrow \infty$. Here, $r = r(u, n)$ is now the unique positive solution of the equation

$$n = \sum_p \frac{pue^{-pr}}{1 + ue^{-pr}}.$$

Before we prove this result, let us first introduce the function $Y(u, s)$ defined to be the Mellin transform of the function $\log(1 + ue^{-x})$. Then the following lemma can be found in [12, Lemma 1]:

Lemma 6. *For any fixed u lying in the cut-plane $\mathbb{C} \setminus (-\infty, -1]$, the function $Y(u, s)$ can be meromorphically continued to the whole s -plane with simple poles at $s = 0, -1, -2, -3, \dots$. Moreover, $Y(u, s)$ satisfies the estimate*

$$|Y(u, \sigma + it)| \ll e^{-(\pi/2 - \varepsilon)|t|}$$

for any $\varepsilon > 0$ and $|t| \rightarrow +\infty$, uniformly as σ and u are restricted to compact sets.

This property follows from the fact that the function $Y(u, s)$ can be written as a product of a polylogarithm and the Gamma function.

Proof. (of Proposition 5) We follow the lines in the proof of Proposition 4, so by the Cauchy theorem we have

$$Q_n(u) = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(int + f(u, r + it)) dt$$

for any $r > 0$. The saddle-point method suggests to choose r a solution of the equation

$$n = -f_1(u, r) = \sum_p \frac{p}{u^{-1}e^{rp} + 1}. \quad (18)$$

The sum on the right hand side is a strictly decreasing function of r tending to 0 when $r \rightarrow +\infty$ and ∞ when $r \rightarrow 0$. Therefore, the solution $r = r(u, n)$ of the equation (18) exists and it is unique for u and n fixed. Also, $r(u, n)$ tends to 0 uniformly in u as $n \rightarrow \infty$. The next step is to split the integral into a central part which is the integral over the interval $[-r^{1+\beta}, r^{1+\beta}]$, where β is a constant such that $1/3 < \beta < 1/2$, and the tails. Let us first evaluate the integral in the center, for which $|t| \leq r^{1+\beta}$ and the function $int + f(u, r + it)$ admits the Taylor expansion

$$int + f(u, r + it) = f(u, r) - f_2(u, r) \frac{t^2}{2} + O(|t^3| \sup_{0 \leq t_0 \leq t} |f_3(u, r + it_0)|).$$

Note that

$$Y(u, 1) = \int_0^{+\infty} \log(1 + ue^{-t}) dt$$

is strictly positive for any value of $u > 0$. By the Mellin transform method and Lemma 6 along with this observation, the function $f_2(u, r)$ is of order $r^{-3}/\log \frac{1}{r}$ and $|f_3(u, r + it)|$ is a $O(r^{-4}/\log \frac{1}{r})$; these estimates are uniform in u . One can use the same technique as in the proof of Proposition 4 to justify the bound on $|f_3(u, r + it)|$ provided that u is close enough to 1 (that is to choose a relatively small δ). Therefore, as in Lemma 4 the integral in the center gives the term we want. The tails are small as a result of the following observation combined with Lemma 2:

$$\begin{aligned} \frac{|Q(u, e^{-(r+it)})|^2}{Q(u, e^{-r})^2} &= \prod_p \left(1 - \frac{2ue^{-pr}(1 - \cos(pt))}{(1 + ue^{-pr})^2} \right) \\ &\leq \exp \left(- \frac{2u}{(1+u)^2} \sum_p e^{-pr}(1 - \cos(pt)) \right). \end{aligned}$$

Finally, from Equation (18) and from the Mellin transform method we derive the asymptotic formula for r

$$r = r(u, n) \sim \sqrt{\frac{2Y(u, 1)}{n \log n}}$$

uniformly in u as $n \rightarrow \infty$. Thus, the result follows by choosing $\beta > 3/7$. \square

Until the end of this section let us use the following abbreviations: $r = r(u, n)$, $r_0 := r(1, n)$, $u = e^{t/\sigma_n}$ and

$$f_{ij}(u, r) = \frac{\partial^i}{\partial^i \tau} \frac{\partial^j}{\partial^j u} f(u, \tau)|_{\tau=r}.$$

Then it follows easily from Proposition 5 that

$$\frac{Q_n(u)}{Q_n(1)} = \exp \left(n(r - r_0) + f(u, r) - f(1, r_0) \right) \left(1 + O \left(\frac{|t|}{\sigma_n} + n^{-1/7} \right) \right).$$

By implicit differentiation we have

$$r - r_0 = -\frac{g_1(r_0)}{f_2(1, r_0)}(u - 1) + O(r_0(u - 1)^2) = O(r_0(u - 1)).$$

Therefore,

$$f(u, r) - f(u, r_0) = f_1(u, r_0)(r - r_0) + f_2(u, r_0)\frac{(r - r_0)^2}{2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}).$$

Also,

$$\begin{aligned} f_1(u, r_0)(r - r_0) &= f_1(1, r_0)(r - r_0) + f_{11}(1, r_0)(u - 1)(r - r_0) \\ &\quad + O(t^3\sqrt{r_0\log\frac{1}{r_0}}) \\ &= -n(r - r_0) + g_1(r_0)(u - 1)(r - r_0) + O(t^3\sqrt{r_0\log\frac{1}{r_0}}), \end{aligned}$$

and for the second term we have

$$f_2(u, r_0)\frac{(r - r_0)^2}{2} = f_2(1, r_0)\frac{(r - r_0)^2}{2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}).$$

Finally,

$$f(u, r_0) - f(1, r_0) = g(r_0)(u - 1) + f_{02}(1, r_0)\frac{(u - 1)^2}{2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}).$$

Thus the function in the exponent can be written as

$$g(r_0)(u - 1) + \left(f_{02}(1, r_0) - \frac{g_1(r_0)^2}{f_2(1, r_0)}\right)\frac{(u - 1)^2}{2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}).$$

On the other hand

$$u - 1 = \frac{t}{\sigma_n} + \frac{t^2}{2\sigma_n^2} + O(t^3/\sigma_n^3)$$

so the exponent becomes

$$\begin{aligned} g(r_0)\frac{t}{\sigma_n} + \left(g(r_0) + f_{02}(1, r_0) - \frac{g_1(r_0)^2}{f_2(1, r_0)}\right)\frac{t^2}{2\sigma_n^2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}) \\ = g(r_0)\frac{t}{\sigma_n} + \left(h(r_0) - \frac{g_1(r_0)^2}{f_2(1, r_0)}\right)\frac{t^2}{2\sigma_n^2} + O(t^3\sqrt{r_0\log\frac{1}{r_0}}) \end{aligned}$$

The error terms in the above expansions can be verified using the same method we used to bound the $|f_3(r + it)|$ in the proof of Proposition 4. Replacing μ_n and σ_n by their respective values in Equation (17), we get the asymptotic formula we expected:

$$M_n(t) = e^{t^2/2}(1 + O((|t| + |t|^3)n^{-1/4+\epsilon} + n^{-1/7})) \quad (19)$$

as $n \rightarrow \infty$, for any $\epsilon > 0$. By Curtiss' theorem [3] the limit distribution is indeed Gaussian.

Remark. The asymptotic formula we get in (19) is very similar to those we find in [12] or [14] and therefore, the following bounds hold for the tails

$$\mathbb{P}\left(\frac{\varpi_n - \mu_n}{\sigma_n} \geq x\right) \leq \begin{cases} e^{-x^2/2} \left(1 + O(1/\log^3 n)\right) & \text{if } 0 \leq x \leq n^{1/12-\epsilon}, \\ e^{-n^{1/12-\epsilon}x/2} \left(1 + O(1/\log^3 n)\right) & \text{if } x \geq n^{1/12-\epsilon}, \end{cases}$$

for any small constant $\epsilon > 0$. Similar bounds hold for

$$\mathbb{P}\left(\frac{\varpi_n - \mu_n}{\sigma_n} \leq -x\right).$$

4. Unrestricted partitions

Recall that unrestricted partitions are those whose parts are allowed to repeat. The appropriate bivariate generating function for the unrestricted case is given by

$$Q(u, z) = \prod_p (1 - uz^p)^{-1}.$$

The infinite product converges only if $|uz| < 1$, unlike the restricted case where we had convergence for $|z| < 1$, for any u restricted in a bounded interval containing 1. As before, we consider the logarithm of the above infinite product

$$f(u, \tau) = \log Q(u, e^{-\tau}) = \sum_p \log(1 - ue^{-p\tau}).$$

Mean and Variance

For the mean and variance, we have formulas rather similar to those for the restricted case. The analogue of the integral representation for the mean is

$$p(n)\mu_n^* = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(int + f(1, r + it))g(r + it)dt \quad (20)$$

for any $r > 0$, where $p(n)$ is the total number of ways of writing n as a sum of primes, and

$$g(\tau) := \sum_p \frac{e^{-p\tau}}{1 - e^{-p\tau}}.$$

For the variance,

$$p(n)(\sigma_n^{*2} + \mu_n^{*2}) = \frac{e^{nr}}{2\pi} \int_{-\pi}^{\pi} \exp(int + f(1, r + it)) (g(r + it)^2 + h(r + it))dt \quad (21)$$

where

$$h(\tau) = \sum_p \frac{e^{-p\tau}}{(1 - e^{-p\tau})^2}.$$

The Mellin transform of the k th derivatives of $f(1, \tau)$ and $g(\tau)$ at $\tau = r$ are given by

$$\mathcal{M}(f_k(1, r), s) = (-1)^k \zeta(s - k + 1) \Gamma(s) D(s - k)$$

and

$$\mathcal{M}(g_k(r), s) = (-1)^k \zeta(s - k) \Gamma(s) D(s - k).$$

So, by Lemma 3 the orders of $f_k(1, r)$ and $g_k(r)$ differ from $r^{-(k+1)}$ only by factors of $\log 1/r$ or $\log \log \frac{1}{r}$, where we take $r = r(n) > 0$ to be the solution of the equation

$$n = \sum_p \frac{pe^{-rp}}{1 - e^{-rp}}.$$

Thus, we only have to repeat the procedures in the previous section to compute the mean. For the variance, the Mellin transform of the function $h(r)$ is given by

$$\mathcal{M}(h(r), s) = \zeta(s - 1)\Gamma(s)D(s)$$

which has a simple pole at $s = 2$ from the zeta function. Therefore, the order of $h(r)$ is r^{-2} which is greater than the contribution from the $g(\tau)^2$ in the integral for the variance. The asymptotic formulas for the mean and variance follow:

Theorem 7. *The mean and variance of the distribution of the number of summands in an unrestricted partition of an integer n into primes satisfy the following asymptotic formulas:*

$$\mu_n^* = \sum_p \frac{e^{-rp}}{1 - e^{-rp}} + O(\log^2 \frac{1}{r})$$

As a function of n ,

$$\mu_n^* = \frac{\sqrt{3}}{\pi} (\log \log n + B_1 - \log 2) \sqrt{n \log n} \left(1 + \frac{\log \log n}{2 \log n} + O\left(\frac{1}{\log n}\right) \right).$$

Likewise,

$$\sigma_n^{*2} = \sum_p \frac{e^{-rp}}{(1 - e^{-rp})^2} + O\left(\frac{\log^2 \frac{1}{r}}{r}\right).$$

As a function of n ,

$$\sigma_n^{*2} = \frac{3D(2)n \log n}{\pi^2} \left(1 + \frac{\log \log n}{\log n} + O\left(\frac{1}{\log n}\right) \right)$$

as $n \rightarrow \infty$.

For comparison, the mean number of summands of a partition into arbitrary parts is

$$\frac{\sqrt{6n}}{2\pi} \left(\log n + 2\gamma - \log(\pi^2/6) \right) + O(\log n),$$

see [11].

Remark. We also have a central limit theorem in the unrestricted case: it is known from [10] that the limit as $n \rightarrow \infty$ of the normalized random variable $\frac{\varpi_n^* - \mu_n^*}{\sigma_n^*}$ has the following moment generating function:

$$M(t) = \prod_p \left(1 - \frac{\tilde{t}}{p} \right)^{-1} e^{-\frac{\tilde{t}}{p}}$$

where $\tilde{t} = t/\sqrt{D(2)}$.

5. Generalization

A natural question one can ask is whether the result remains true for powers of primes or more generally for polynomials $f(p)$ of primes. But one needs to be careful here since, for example, $p^2 + p$ is always even for any prime p so we need to impose some additional conditions on the polynomial. From the result of [17] and a slight modification of our proof of Lemma 2 we can get

Lemma 8. *Let $f(x)$ be a strictly increasing polynomial which takes only integral values for integer x and has the property that for every prime p there is a positive integer x such that $p \nmid xf(x)$. For any constant $1/3 < c < 1/2$ and $r^{1+c} \leq |y| \leq \pi$ we have the inequality*

$$\sum_p e^{-f(p)r} (1 - \cos f(p)y) \geq c' \log^2 \frac{1}{r}$$

for an absolute constant $c' > 0$ as $r \rightarrow 0^+$.

The associated Dirichlet series is closely related to the Dirichlet series of primes. Suppose that the dominant term in our polynomial $f(x)$ is of the form ax^d where a is a positive integer and d is the degree of $f(x)$. Then

$$D_f(s) - a^{-s} D(ds) = a^{-s} \sum_p \left(\frac{a^s p^{ds} - (f(p))^s}{p^{ds} (f(p))^s} \right).$$

The series on the right hand side is absolutely convergent for $\text{Re}(s) > 1/(2d)$. Therefore our method applies, and the limit distribution of the number of summands in partitions of n into distinct primes is Gaussian with mean and variance

$$\mu_n \sim C_1(a, d) \left(\frac{n}{\log^d n} \right)^{\frac{1}{1+d}} \quad \text{and} \quad \sigma_n^2 \sim C_2(a, d) \left(\frac{n}{\log^d n} \right)^{\frac{1}{1+d}},$$

where $C_i(a, d)$ can be determined explicitly. As for unrestricted partitions, the mean and variance follow the asymptotic formulas:

$$\mu_n^* \sim C_1'(a, d) (n \log n)^{\frac{d}{1+d}} \quad \text{and} \quad \sigma_n^{*2} \sim C_2'(a, d) (n \log n)^{\frac{2d}{1+d}}$$

when $d \geq 2$, again the $C_i'(a, d)$ can be determined explicitly.

Acknowledgments

Many thanks to Stephan Wagner for his comments and remarks on an earlier draft of this paper. The author also wants to thank the German Academic Exchange Service (DAAD) and the African Institute for Mathematical Sciences (AIMS) for their financial support.

References

- [1] Tom M. Apostol. *Introduction to analytic number theory*. Springer-Verlag, New York, 1976. Undergraduate Texts in Mathematics.
- [2] Charlotte Brennan, Arnold Knopfmacher, and Stephan Wagner. The distribution of ascents of size d or more in partitions of n . *Combin. Probab. Comput.*, 17(4):495–509, 2008.
- [3] J. H. Curtiss. A note on the theory of moment generating functions. *Ann. Math. Statistics*, 13:430–433, 1942.
- [4] Paul Erdős and Joseph Lehner. The distribution of the number of summands in the partitions of a positive integer. *Duke Math. J.*, 8:335–345, 1941.
- [5] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
- [6] Kevin Ford. Vinogradov’s integral and bounds for the Riemann zeta function. *Proc. London Math. Soc. (3)*, 85(3):565–633, 2002.
- [7] William M. Y. Goh and Eric Schmutz. The number of distinct part sizes in a random integer partition. *J. Combin. Theory Ser. A*, 69(1):149–158, 1995.
- [8] G. H. Hardy and S. Ramanujan. Asymptotic formulæ for the distribution of integers of various types [Proc. London Math. Soc. (2) **16** (1917), 112–132]. In *Collected papers of Srinivasa Ramanujan*, pages 245–261. AMS Chelsea Publ., Providence, RI, 2000.
- [9] G. H. Hardy and S. Ramanujan. Asymptotic formulæ in combinatory analysis [Proc. London Math. Soc. (2) **17** (1918), 75–115]. In *Collected papers of Srinivasa Ramanujan*, pages 276–309. AMS Chelsea Publ., Providence, RI, 2000.
- [10] C. B. Haselgrove and H. N. V. Temperley. Asymptotic formulae in the theory of partitions. *Proc. Cambridge Philos. Soc.*, 50:225–241, 1954.
- [11] K. Husimi. Partitio numerorum as occurring in a problem of nuclear physics. *Proceedings of the Physico-Mathematical Society of Japan.*, 20:912–925, 1938.
- [12] Hsien-Kuei Hwang. Limit theorems for the number of summands in integer partitions. *J. Combin. Theory Ser. A*, 96(1):89–126, 2001.
- [13] D. V. Lee. The asymptotic distribution of the number of summands in unrestricted Λ -partitions. *Acta Arith.*, 65(1):29–43, 1993.
- [14] Manfred Madritsch and Stephan Wagner. A central limit theorem for integer partitions. *Monatsh. Math.*, 161(1):85–114, 2010.
- [15] L. B. Richmond. Some general problems on the number of parts in partitions. *Acta Arith.*, 66(4):297–313, 1994.
- [16] L. Bruce Richmond. The moments of partitions. II. *Acta Arith.*, 28(3):229–243, 1975/76.
- [17] K. F. Roth and G. Szekeres. Some asymptotic formulae in the theory of partitions. *Quart. J. Math., Oxford Ser. (2)*, 5:241–259, 1954.
- [18] R. C. Vaughan. On the number of partitions into primes. *Ramanujan J.*, 15(1):109–121, 2008.

Dimbinaina Ralaivaosaona
Stellenbosch University
Department of Mathematical Sciences
Mathematics Division
Private Bag X1
Matieland 7602, South Africa
e-mail: naina@sun.ac.za