

Counting integer partitions with the method of maximum entropy

Joint work with Marcus Michelen and Will Perkins

Gwen McKinley

May 7, 2021

University of California, San Diego

- **Big idea:** a different technique (“principle of maximum entropy”) allows us to approach an old problem (enumerating integer partitions) with new intuition and a more powerful/flexible solution.
- Sketch of the method for a classical example (Hardy-Ramanujan asymptotic partition formula)
- Variations on the classical problem
- Our result
- Time permitting, a few ideas from the proof of one part (Local CLT)

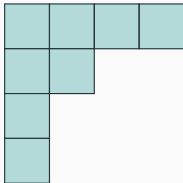
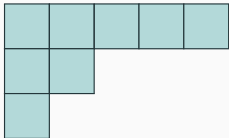
Integer partitions

Definition

A *partition* of a positive integer n is a representation of n as an unordered sum of positive integers.

Example:

- $5 + 2 + 1$ and $4 + 2 + 1 + 1$ are both partitions of 8.
- $5 + 2 + 1$ and $1 + 2 + 5$ are the *same* partition of 8.



Integer partitions

Definition

A *partition* of a positive integer n is a representation of n as an unordered sum of positive integers.

Question: How many different partitions of n are there? Write $P(n)$ for the set of partitions of n , and $p(n)$ for the number.

E.g. $p(4) = 5$:

- 4
- $3 + 1$
- $2 + 2$
- $2 + 1 + 1$
- $1 + 1 + 1 + 1$

Integer partitions

Definition

A *partition* of a positive integer n is a representation of n as an unordered sum of positive integers.

Problem: How many different partitions of n are there? Write $P(n)$ for the set of partitions of n , and $p(n)$ for the number.

For the first few values, $p(n)$ is

1, 2, 3, 5, 7, 11, 15, 22, 30, 42, 56, 77, 101, 135, 176, 231, 297

In general, **very hard!** No closed form known.

Problem 2.0: Find asymptotic behavior of $p(n)$ as $n \rightarrow \infty$.

Theorem (Hardy and Ramanujan, 1918)

$$p(n) = \frac{1 + o(1)}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}.$$

Counting partitions

Theorem (Hardy and Ramanujan, 1918)

$$p(n) = \frac{1 + o(1)}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}.$$

Question: Intuitive explanation? Even just for the exponent?

- Original proof: *circle method*.
- Extract $p(n)$ from generating function with Cauchy's residue formula. $\Rightarrow$ need to evaluate nasty complex integral.
- Our idea: *principle of maximum entropy*.

Warning! Fuzzy math ahead.

Maximum entropy

- **Information-theoretic approach:** try to understand partitions of n by looking at some probability distribution on partitions of *any* integer.
- Which distribution to choose?
- Jaynes' principle of maximum entropy: “best” distribution has *maximum entropy* among all distributions that give a partition of n in expectation.
- Best how?



Maximum entropy

Fact

On a finite set S , the uniform distribution has the largest entropy of any distribution: $\log |S|$.

If X is uniform on $P(n)$, we have $p(n) = e^{H(X)}$. *Not any easier!*

What about an *almost* uniform distribution?

Goal: find a distribution X (on partitions of *any* integer) where we *can* compute its entropy, and with $p(n) \approx e^{H(X)}$.

Maximum entropy

Goal: find a distribution X (on partitions of *any* integer) where we *can* compute its entropy, and with $p(n) \approx e^{H(X)}$.

Try Jaynes' principle of maximum entropy. Here, it says:

Find the maximum entropy distribution $X = (X_1, X_2, \dots)$ on $\mathbb{N} \times \mathbb{N} \times \dots$ (where X_k = multiplicity of k) subject to

$$\mathbb{E} \left[\sum_{k \geq 1} k \cdot X_k \right] = n.$$

Magic fact: For the entropy-maximizing distribution X , we have

$$p(n) = e^{H(X)} \cdot \Pr[X \in P(n)].$$

Maximum entropy

Magic Lemma (general)

$\Omega =$ countably infinite set

$f =$ function $\Omega \rightarrow \mathbb{R}_+^d$

$\mathbf{N} \in \mathbb{R}^d$ “constraint vector”

$S_{\mathbf{N}} = \{\omega \in \Omega : f(\omega) = \mathbf{N}\}$

partitions of *any* integer

sum of parts

n

$P(n)$

Suppose that $S_{\mathbf{N}}$ is finite and non-empty, and that there exists a distribution μ on Ω **of the right form** with $\mathbb{E}_{\mu} f = \mathbf{N}$. Then μ is the maximum entropy distribution on Ω so that $\mathbb{E}_{\mu} f = \mathbf{N}$, and

$$|S_{\mathbf{N}}| = e^{H(\mu)} \cdot \mu(S_{\mathbf{N}}).$$

Maximum entropy

Magic Lemma

Suppose that $S_{\mathbf{N}}$ is finite and non-empty, and that there exists a distribution μ on Ω **of the right form** with $\mathbb{E}_{\mu} f = \mathbf{N}$. Then μ is the maximum entropy distribution on Ω with $\mathbb{E}_{\mu} f = \mathbf{N}$, and

$$|S_{\mathbf{N}}| = e^{H(\mu)} \cdot \mu(S_{\mathbf{N}}).$$

- **Note:** there's not always a maximum entropy distribution satisfying the constraint. But...
- **Sufficient:** “right form” is $\mu(\omega) = \frac{1}{Z} e^{-\mathbf{b} \cdot \mathbf{f}(\omega)}$ for some $\mathbf{b} \in \mathbb{R}^d$ with $Z := \sum_{\omega \in \Omega} e^{-\mathbf{b} \cdot \mathbf{f}(\omega)} < \infty$.

Maximum entropy

Recap: Want to find max entropy $X = (X_1, X_2, \dots)$ subject to $\mathbb{E}[\sum_{k \geq 1} k \cdot X_k] = n$. Then by “magic fact,” have

$$p(n) = e^{H(X)} \cdot \Pr[X \in P(n)].$$

Start with any distribution $(Y_1, Y_2, \dots)$.

- **Fact 1:** “Decoupling” the marginals Y_k increases entropy.
- **Fact 2:** Replacing any Y_k with a geometric r.v. with mean $\mu_k = \mathbb{E}[Y_k]$ increases entropy.

$\Rightarrow$ Max entropy $(X_1, X_2, \dots)$ has independent geometric X_k 's.
Just need the right sequence of means $(\mu_1, \mu_2, \dots)$.

Maximum entropy

New problem: Find right sequence of means $(\mu_1, \mu_2, \dots)$ to maximize the entropy of the corresponding distribution $(X_1, X_2, \dots)$ of independent geometric random variables, subject to $\sum_{k \geq 1} k \cdot \mu_k = n$.

Fact

A geometric r.v. with mean μ has entropy

$$G(\mu) := (\mu + 1) \log(\mu + 1) - \mu \log \mu.$$

Corresponds to a discrete optimization problem:

$$\text{Maximize } \sum_{k \geq 1} G(\mu_k), \quad \text{subject to } \sum_{k \geq 1} k \cdot \mu_k = n.$$

Maximum entropy

$$\begin{aligned} &\text{Maximize} && \sum_{k \geq 1} G(\mu_k), \\ &\text{subject to} && \sum_{k \geq 1} k \cdot \mu_k = n. \end{aligned}$$

Rescale by writing $m(x) := \mu_{x\sqrt{n}}$, “massage” the sums algebraically, and interpret them as Riemann sums. Then as $n \rightarrow \infty$, approximately a continuous optimization problem:

$$\begin{aligned} &\text{Maximize} && \sqrt{n} \cdot \int_0^\infty G(m(x)) \, dx, \\ &\text{subject to} && \int_0^\infty x \cdot m(x) \, dx = 1. \end{aligned}$$

Maximum entropy

$$\begin{aligned} &\text{Maximize} && \sum_{k \geq 1} G(\mu_k), \\ &\text{subject to} && \sum_{k \geq 1} k \cdot \mu_k = n. \end{aligned}$$

Rescale by writing $m(x) := \mu_{x\sqrt{n}}$, “massage” the sums algebraically, and interpret them as Riemann sums. Then as $n \rightarrow \infty$, approximately a continuous optimization problem:

$$\begin{aligned} &\text{Maximize} && \sqrt{n} \cdot \int_0^\infty G(m(x)) \, dx, \\ &\text{subject to} && \int_0^\infty x \cdot m(x) \, dx = 1. \end{aligned}$$

Maximum entropy

$$\begin{aligned} &\text{Maximize} && \int_0^\infty G(m(x)) dx, \\ &\text{subject to} && \int_0^\infty x \cdot m(x) dx = 1. \end{aligned}$$

Pretty easy! Can use Lagrange multipliers (continuous “calculus of variations” version). Solve to find the optimizer, $m^*(x) = \frac{1}{e^{\frac{\pi}{\sqrt{6}}x} - 1}$, and plug in to get our final answer:

$$H(X) = \sum_{k \geq 1} G(\mu_k) \approx \sqrt{n} \cdot \int_0^\infty G(m^*(x)) dx = \sqrt{n} \cdot \pi \sqrt{\frac{2}{3}}.$$

Look familiar? :)

Maximum entropy

Recap: Wanted to find max entropy distribution X on partitions with expected sum n .

“**Magic fact**” $\Rightarrow p(n) = e^{H(X)} \cdot \Pr[X \in P(n)]$

We’ve approximated $e^{H(X)} \approx e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}$ by solving a continuous optimization problem. *Exponential term in Hardy-Ramanujan!*

Remaining questions:

- Error from $\sum \rightarrow \int$? $\frac{1}{\sqrt[4]{24}n^{1/4}}$
- What is $\Pr[X \in P(n)]$? Probability that $\sum_{k \geq 1} k \cdot X_k$ hits its mean of n . *Prove a (local) central limit theorem.* $\frac{1}{2\sqrt[4]{6}n^{3/4}}$

Multiply to get $\frac{1+o(1)}{4\sqrt[4]{3}n} e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}$. *Hardy-Ramanujan!!*

⚡ **CONTINUE**
SAVE

WARNING
THIS GAME IS REALLY
DIFFICULT.

Generalizations & related work

Asymptotic count known for many “flavors” of partitions of n , e.g.,

- $\leq k$ parts (Szekeres, 1953 + others)
- $\leq k$ parts, difference $\geq d$ between parts (Romik, 2005)
- parts are k^{th} powers (Wright, 1934 + others)
- $\leq k$ parts, each $\leq \ell$, “ q -binomial coefficients” (e.g. Melczer, Panova, and Pemantle, 2019, and Jiang and Wang, 2019)

Also, many papers studying the structure of a “typical” partition (e.g. Fristedt, 1997)

Methods including:

- Circle method (many)
- Use results about “typical” partitions + prove a local central limit theorem (e.g. Romik, 2005)
- “Physics stuff” (e.g. Tran, Murthy, and Badhuri, 2003)
- Large deviations (Melczer, Panova, and Pemantle, 2019)

No free lunch – usually some messy integrals.

Related: “counting via maximum entropy” e.g. for counting lattice points in polytopes (Barvinok and Hartigan, 2010).

Our result

Can use the “maximum entropy” approach for any of these:
becomes a constrained optimization problem with more constraints.
But many a slip 'twixt the cup and the lip... (especially: local CLT)

As a “proof of concept”, we’ll count the following partitions:

Definition

Given a finite index set $J \subset \mathbb{N}$, and a vector of positive integers $\mathbf{N} = (N_j)_{j \in J}$, we say that a partition P has *profile* $\mathbf{N}$ if

$$\sum_{x \in P} x^j = N_j \text{ for all } j \in J.$$

Definition

Given a finite index set $J \subset \mathbb{N}$, and a vector of positive integers $\mathbf{N} = (N_j)_{j \in J}$, we say that a partition P has *profile* $\mathbf{N}$ if

$$\sum_{x \in P} x^j = N_j \text{ for all } j \in J.$$

Write $p(\mathbf{N})$ for the number of such partitions.

- “Unrestricted” partitions ($J = \{1\}$)
- Partitions with fixed $\#$ of parts ($J = \{0, 1\}$)
- Partitions of n into k^{th} powers ($J = \{k\}$ and $n = N_k$)

Our result

Notation: For any index set J , and any $\alpha \in \mathbb{R}_+^{|J|}$, write $N = (N_j)_{j \in J} = (\lfloor \alpha_j n^{(j+1)/2} \rfloor)_{j \in J}$.

Assume: there exists $\beta \in \mathbb{R}^J$ that is a solution to the system of equations

$$\int_0^\infty \frac{x^j}{\exp\left(\sum_{j \in J} \beta_j x^j\right) - 1} dx = \alpha_j \quad \text{for } j \in J.$$

Then define:

$$\begin{aligned} M(\alpha) = \text{maximum of } & \int_0^\infty G(m(x)) dx, \\ \text{subject to } & \int_0^\infty x^j \cdot m(x) dx = \alpha_j, \quad \text{for all } j \in J. \end{aligned}$$

Main Theorem (M., Michelen, and Perkins, 2020?)

For any index set J , and any $\alpha \in \mathbb{R}_+^{|J|}$,

$$p(\mathbf{N}) = (1 + o(1)) \frac{e^{M(\alpha)\sqrt{n}}}{c_1(\alpha) \cdot n^{c_2(\alpha)}}$$

if $\mathbf{N}$ is “feasible”.

- $M(\alpha)$ = solution to continuous optimization problem (constant), and $M(\alpha)\sqrt{n}$ = entropy of max entropy distribution, after approximating $\sum \rightarrow \int$.
- c_1, c_2 constants.
- Other terms: error from $\sum \rightarrow \int$, and probability that max entropy distribution hits $P(\mathbf{N})$. (*Local CLT – rest of talk*)

Local CLT (M., Michelen, and Perkins, 2020?)

$X = (X_1, X_2, \dots)$ a joint distribution of independent geometric r.v.s with appropriate parameters. Write $\mathbf{N}_X = (\sum_{k \geq 1} k^j X_k)_{j \in J}$ (the profile of X). Then for any possible profile $\mathbf{a} \in \mathbb{N}^J$,

$$\Pr(\mathbf{N}_X = \mathbf{a}) \approx \mathbb{1}_{\mathbf{a} \text{ is "feasible"}} \left(\frac{\# \text{ integer-valued polys.}}{\text{in some region}} \right) \cdot (\text{PDF of Gaussian})$$

- Many impossible profiles $\mathbf{a}$, e.g. $\mathbf{a}_1 = (\text{even})$ and $\mathbf{a}_2 = (\text{odd})$.
- $\Rightarrow \Pr(\mathbf{N}_X = \mathbf{a}) = 0$ in many places
- $\Rightarrow$ probability mass “piles up” on remaining points.
- Extra factor on remaining points (“feasible” points).

Local CLT (M., Michelen, and Perkins, 2020?)

$X = (X_1, X_2, \dots)$ a joint distribution of independent geometric r.v.s with appropriate parameters. $\mathbf{N}_X = (\sum_{k \geq 1} k^j X_k)_{j \in J}$. Then

$$\Pr(\mathbf{N}_X = \mathbf{a}) \approx \mathbb{1}_{\mathbf{a} \text{ is "feasible"}} \left(\# \begin{smallmatrix} \text{integer-valued polys.} \\ \text{in some region} \end{smallmatrix} \right) \cdot (\text{PDF of Gaussian})$$

Proof ideas:

- Want to understand PMF of $\mathbf{N}_X$, and know that $\mathbf{N}_X$ is defined in terms of sums of independent geometric r.v.s.
- Work with the characteristic functions of the X_k 's.
- Characteristic function = Fourier transform of PMF, so to extract PMF from characteristic function: Fourier inversion.

Local CLT (M., Michelen, and Perkins, 2020?)

$$\Pr(\mathbf{N}_X = \mathbf{a}) \approx \mathbb{1}_{\mathbf{a} \text{ is "feasible"}} \left(\# \begin{smallmatrix} \text{integer-valued polys.} \\ \text{in some region} \end{smallmatrix} \right) \cdot (\text{PDF of Gaussian})$$

Proof ideas:

- Fourier inversion gives $\Pr(\mathbf{N}_X = \mathbf{a})$ as a nasty complex integral in terms of characteristic functions.
- Throw away regions that “obviously” don’t contribute much.
- **Green-Tao (2012):** this leaves us with a neighborhood around the coefficients of each integer-valued polynomial.
- On each neighborhood, approximate with a Gaussian.

Recap:

- Max entropy approach gives # of partitions (with restrictions allowed) as $e^{H(X)} \cdot \Pr[X \in P]$, where X = max entropy distribution.
- $e^{H(X)}$ fairly easy to find! *Leading constant in $H(X)$ given by a continuous optimization problem.*
- Still no free lunch though: for lower-order terms, have to approximate $\sum \rightarrow \int$ error, and (more difficult) to find $\Pr[X \in P]$, have to deal with nasty complex integral by proving local CLT.

Thank you!