Counting integer partitions with the method of maximum entropy

Joint work with Marcus Michelen and Will Perkins

Gwen McKinley

October 1, 2020

University of California, San Diego

Outline

- **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)

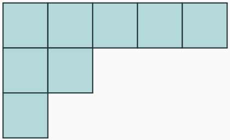
Integer partitions

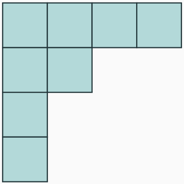
Definition

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 $\boldsymbol{P(n)}$ for the set of partitions of n, and $\boldsymbol{p(n)}$ 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 $\boldsymbol{P(n)}$ for the set of partitions of n, and $\boldsymbol{p(n)}$ 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.

Counting partitions

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

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

- **Probabilistic 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?


LET'S TALK ABOUT THAT

Maximum entropy

Definition

Given a discrete random variable X, the *entropy* of X is

$$H(X) := \sum_x \Pr(X = x) \log \left(\frac{1}{\Pr(X = x)} \right).$$

Measures the amount of "randomness" or "information" in X.

Fact

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

So $|S| = e^{H(X)}$ if X is uniform. *Not any easier!*

Maximum entropy

Fact: If X is uniform on $P(n)$, we have $\boldsymbol{p(n) = e^{H(X)}}$.

Entropy of uniform distribution too hard to compute :(
But what about an *almost* uniform distribution?

Hope: maybe we can find a distribution X (on partitions of *any* integer) that's...

- constant(ish) on $P(n)$,
- fairly concentrated on $P(n)$,
- and where we *can* compute its entropy.

Then maybe $\boldsymbol{p(n) \approx e^{H(X)}}$. *Very sketchy.*

Maximum entropy

Idea: Want an "almost uniform" distribution X on partitions of any integer where we can compute $H(X)$. Hope that $\boldsymbol{p(n) \approx e^{H(X)}}$.

What's the "best" distribution? 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

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

Maximum entropy

Problem: Find max entropy $X = (X_1, X_2, \dots)$ subject to $\mathbb{E}\left[\sum_{k \geq 1} \boldsymbol{k \cdot X_k}\right] = \boldsymbol{n}$.

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:

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

Maximum entropy

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

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

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

Maximum entropy

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

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

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

Maximum entropy

$$\text{Maximize} \quad \int_0^\infty \boldsymbol{G(m(x))\,dx},$$
$$\text{subject to} \quad \int_0^\infty \boldsymbol{x \cdot m(x)\,dx} = 1.$$

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. Hoped that $\boldsymbol{p(n) \approx e^{H(X)}}$.

We've approximated $e^{H(X)} \approx \boldsymbol{e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}}$. *Correct exponential term in Hardy-Ramanujan!*

Method: Solve continuous optimization problem (approximates $\sum$ with $\int$).

Can we make this less sketchy?

Maximum entropy

Wanted to find max entropy distribution X on partitions with expected sum n. Hoped that $\boldsymbol{p(n) \approx e^{H(X)}}$.

Question: *How close to the truth is this assumption?*

Answer: For the maximizing distribution X, we have

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

"Reason": compute directly from distribution.

Maximum entropy

Magic fact

Let $X = (X_1, X_2, ...,)$ be given by a probability distribution satisfying some set of constraints in expectation, and where we've specified the support of the X_k's (must be discrete). For a wide variety of such constraints, if X is the entropy maximizing distribution, we will have:

$$\binom{\#\text{ vectors satisfying}}{\text{the constraints}} = \Pr[\boldsymbol{X} \text{ satisfies constraints}] \cdot \boldsymbol{e}^{\boldsymbol{H}(\boldsymbol{X})}.$$

- **"Just do it"** – max entropy distribution will always have independent X_k's of a specified type.
- Use constraints + Lagrange multipliers to pin down parameters, then compute directly from distribution.

Maximum entropy

Recap: Wanted to find max entropy distribution X on partitions with expected sum n. Initially hoped that $p(n) \approx e^{H(X)}$.

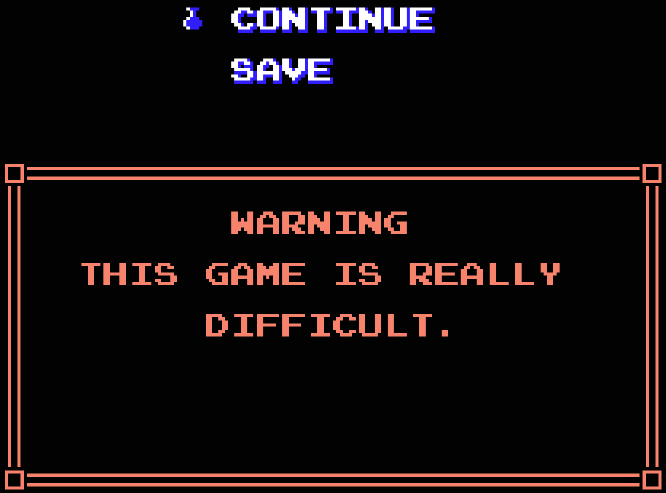
We've approximated $e^{H(X)} \approx \boldsymbol{e^{\pi\sqrt{\frac{2}{3}}\sqrt{n}}}$.

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

Remaining questions:

- Error from $\sum \to \int$? $\frac{1}{\sqrt[4]{24}\boldsymbol{n}^{1/4}}$
- 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.* $\frac{1}{2\sqrt[4]{6}\boldsymbol{n}^{3/4}}$

Multiply to get $\frac{1+\boldsymbol{o}(1)}{4\sqrt{3}\boldsymbol{n}}\boldsymbol{e}^{\boldsymbol{\pi}\sqrt{\frac{2}{3}}\sqrt{\boldsymbol{n}}}$. *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.,

- $\leq \boldsymbol{k}$ parts (Szekeres, 1953 + others)
- $\leq \boldsymbol{k}$ parts, difference $\geq \boldsymbol{d}$ between parts (Romik, 2005)
- parts are $\boldsymbol{k}^{\text{th}}$ powers (Wright, 1934 + others)
- $\leq \boldsymbol{k}$ 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)

Generalizations & related work

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 $\boldsymbol{N} = (N_j)_{j \in J}$, we say that a partition P has *profile* $\boldsymbol{N}$ if

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

Our result

Definition

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

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

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

- "Unrestricted" partitions ($\boldsymbol{J} = \{1\}$)
- Partitions with fixed # of parts ($\boldsymbol{J} = \{0, 1\}$)
- Partitions of n into k^{th} powers ($\boldsymbol{J} = \{\boldsymbol{k}\}$ and $\boldsymbol{n} = \boldsymbol{N}_{\boldsymbol{k}}$)

Our result

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

$$M(\boldsymbol{\beta}) = \text{maximum of} \quad \int_0^\infty G(m(x))\,dx,$$

$$\text{subject to} \quad \int_0^\infty x^j \cdot m(x)\,dx = \beta_j, \quad \text{for all } j \in J.$$

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

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

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

if $\boldsymbol{N}$ is "feasible".

Our result

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

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

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

if $\boldsymbol{N}$ is "feasible".

- $\boldsymbol{M(\beta)\sqrt{n}}$ = entropy of max entropy distribution, after approximating $\sum \to \int$. $\boldsymbol{M(\beta)}$ = solution to continuous optimization problem (constant)
- $\boldsymbol{c}_1, \boldsymbol{c}_2$ constants.
- Other terms: error from $\sum \to \int$, and probability that max entropy distribution hits $P(\boldsymbol{N})$. *(Local CLT – rest of talk)*

Local CLT

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

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

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

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

Local CLT

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

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

$$\Pr(\boldsymbol{N}_X = \boldsymbol{a}) \approx \mathbb{1}_{\boldsymbol{a} \text{ is ``feasible''}} (\# \substack{\text{integer-valued polys.} \\ \text{in some region}}) \cdot (\text{PDF of Gaussian})$$

Proof ideas:

- Want to understand PMF of $\boldsymbol{N}_X$, 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

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

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

Proof ideas:

- Fourier inversion gives $\Pr(\boldsymbol{N}_X = \boldsymbol{a})$ 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.

Overall recap

Recap:

- Max entropy approach gives # of partitions (with restrictions allowed) as $\Pr[\boldsymbol{X} \in \boldsymbol{P}] \cdot \boldsymbol{e}^{\boldsymbol{H}(\boldsymbol{X})}$, 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 $\sum \to \int$ error, and (more difficult) to find $\Pr[\boldsymbol{X} \in \boldsymbol{P}]$, have to deal with nasty complex integral by proving local CLT.

Thank you!