An Useless Method to Measure the von Neumann Entropy

Charles Wang

Somewhere in Cambridge, MA

A Historical Note:

This paper was written on a lark based on a calculation that, while potentially interesting, looks to have no application whatsoever. It was suggested to me that I write the results up. It was not specified, however, that the paper was not to be stuffed full of as many bad jokes as I could come up with. As a result there is a formula or two for the entropy, a discussion of the optimal way to measure entropies with a fixed number of copies of a state, an analysis of asymptotics, and a developing feud with the replica trick. There is also flying papers, footnotes, rhyming, tabloids, cheesemaking, a sufficiently large N, and additionally a developing feud with the replica trick. The following is most likely not useful. But, perhaps, they may be entertaining.

I have endeavored to write this ‘paper’ in as ludicrous manner as possible so that what it lacks in scientific merit, it may lack in comedic merit as well. The following may well be confusing. If you are not confused, please let me know; do teach me what is going on.

1 Replica Blues

The start of this dreadful affair comes from two sources. The first is the question of how to actually, you know, *measure* the von Neumann entropy. The second is that lovely little piece of black magic known as the replica trick.

Both of these revolve around a single quantity, the von Neumann entropy, which equals

$$S = -\mathrm{Tr}\,(\rho \log \rho). \tag{1.1}$$

A simple enough expression, but it seems to like to play hard to get when it actually needs to be calculated. This can be resolved by using the old physicist’s trick of calculating something easier and hoping the problem goes away. The something easier in this case is the Renyi entropy,

$$S_\alpha = \frac{1}{1-\alpha} \log \mathrm{Tr}\,(\rho^\alpha). \tag{1.2}$$

When α is an integer, this often is not so bad and can often be solved by clever staring at path integrals.

The replica trick is based on the following principle: knowing the Renyi entropy to a high degree of accuracy for integer $n > 1$ will tell you the von Neumann entropy, the limit of the Renyi entropy as n approaches 1, to a high degree of accuracy. This may not seem obvious at first glance, but as a matter of fact it is entirely false[1].

Consider the following situation. You flip a weighted coin. Most of the time, with probability $1-\kappa$, the coin lands heads and nothing happens. If the coin lands tails, Daniel Harlow throws a printout of his latest paper at you[2]. What’s the entropy? This comes out to

$$S = -(1-\kappa)\ln(1-\kappa) - \sum_i \kappa p_i \ln(\kappa p_i) = -\kappa \ln(\kappa) - (1-\kappa)\ln(1-\kappa) + \kappa S_H, \tag{1.3}$$

where the sum is over the microstates of the paper, and S_H is the entropy of the text radiated by Harlow[3].

[1] For those who may have noticed the ‘high degree of accuracy’, even with perfect accuracy Carlson’s theorem won’t save you. In an infinite dimensional Hilbert space, the Renyi entropy is *never* an entire function as it is not defined for $Re(n) < 0$. It can easily become non-analytic the moment n hits 1 as well.

[2] This is liable to cause significant blunt force trauma, leading to an undesired increase in entropy. Attempt only with professional supervision

[3] Most of this radiation comes out in discrete quanta known as ‘letters’, while smaller amounts of the

As for the Renyi entropies, one has

$$e^{-(n-1)S_n} = (1-\kappa)^n + \sum_i \kappa^n p_i^n = 1 + O(n\kappa) + O(\kappa^n e^{-nS_H}). \tag{1.4}$$

Therefore, the entropy is zero at leading order and the dependence on S_H is exponentially suppressed. The von Neumann entropy would beg to differ.

One[4] may want to find an estimate for the von Neumann entropy using the Renyis that doesn't involve such Mafia dealings. There is such a formula. It is entirely useless. $\log(\rho)$ can be Taylor expanded around $\rho = 1$ and all eigenvalues of ρ are in the radius of convergence. Plugging this into the formula for the von Neumann entropy gives

$$S = -\mathrm{Tr}\,(\rho \log(\rho)) = \sum_{k=1}^{\infty} \frac{1}{k} \mathrm{Tr}\,(\rho(1-\rho)^k). \tag{1.5}$$

$\mathrm{Tr}\,(\rho(1-\rho)^k)$ is positive and monotonically decreasing in k. Effectively, it computes the amount of probability under progressively tightening cutoffs. The N'th partial sum can be expressed as the expectation value of an operator on N copies of the state ρ. For N sufficiently large[5], this gives an arbitrarily good approximation to the entropy. Therefore, the partial sums of the series monotonically increase, converging on the correct entropy. Hooray! We have our formula and we can now merrily go off to do our calculations[6].

How many terms do we need? The k'th term is less than $1/k$, so this gives at least $\exp(S)$ terms in order to get close to the right answer. Wups.

Well that's not necessarily too bad as some of the corrections are nonpeturbative anyways. Each of the $1-\rho$ terms expands in binomial coefficients. None of these cancel. The order? Exponential in N, the number of terms. Doubly exponential in S, that is. No dice[7].

This approach might actually be salvageable. However, it will turn out to crucially depend on solving the worst possible limit of the replica trick and on understanding the results of a mathematics paper that I have, as of yet, been completely unable to comprehend.

radiation is in the form of more exotic states. For an overview of such phenomena, see *Physical Review Letters vol. 1-26.*

[4]Who is this 'one' person anyways?

[5]For an N sufficiently large for most purposes, see Appendix A.

[6]This version of the entropy estimate is actually slightly different from the version derived and analyzed later.

[7]We'll be needing them later.

2 The Hunting of the $S_N$invars

Now who came up with this 'von Neumann entropy' anyways[8]? How do you figure what the damn thing is for anything? Since rhetorical questions are meant to be answered, here is a canonical procedure for measuring the entropy, at least classically. The quantum version will be coming soon.

Suppose you want to measure the entropy of a loaded die. There's only one way to go about it: Take a bucket full of identical copies of them[9] and then pour the bucket on the floor[10]. Count the number of times each face lands face up, calculate the Shannon entropy, and there is your estimate. Note how the number of dice needed is exponential in the entropy. There is no better that you can do: for $N \ll \exp(S/2)$ and probability not too unevenly distributed, each outcome will typically be different. When faced with N distinct outcomes there's nothing much better to do than just guess $\log(N)$, which will be quite a ways off in general.

And now, quantum mechanics. The expectation of an operator can be measured to any degree of accuracy through repetition. That's the whole point of that expectation value thing. Therefore, measurement actually lives in the space of operators on the tensor product of N copies of the state. Note how each measurement kills the quantum (and classical) superposition of the state, so a new copy must be sacrificed for the next measurement[11].

For entropies, the optimal operators turn out be highly constrained: they are sums of projectors onto irreducible representations of the symmetric group. There is a $SU(N)$ symmetry on the Hilbert space along with a S_N permutation group so in order to measure the N'th Renyi entropy S_N on the N copies-

Hold it.

There are simply *too many N's.* Some symbol reshuffling is in order to resolve the problem before things get worse.
This, of course, can only be accomplished in verse.

- n: The number of tensored copies of the state upon which you run your experiments.
- N: The dimension of the Hilbert space containing the state and its variants.
- M: An arbitrarily large number from later whose meaning will be seen.
- N: An English letter. (See P.R.L. vol. 14.)
- k: The order of the Renyi Entropy which was an n before.

[8]Claude Shannon, in fact, since the next bit is going to be some classical information theory.
[9]Useful if you want to get beaten up in a dark alley by some nice men from the Casino.
[10]You can also roll one die many times, but where is the fun in that?
[11]Campaigns for the humane treatment of quantum states have thusfar not been successful.

- H_k: The Renyi entropy of order k: A family including the von Neumann and more.
- S_n: The symmetric group on n elements whose action preserves the state.
- $\mathbb{N}$: The natural numbers, such as ninety-eight.
- $U(N)$: The unitary group on the Hilbert space which keeps H_k the same.
- H: The von Neumann entropy whose value is our aim.

Got it?

So what operators should you consider? There *are* a whole lot of them, after all. Symmetry is a big help. There is a $U(N)$ symmetry acting upon the Hilbert space and there is a S_n permutation group shuffling the states around. Transforming an density matrix under $U(N)$ doesn't change the entropy and gives another density matrix just as good as any other. Therefore, conjugating any operator giving an estimate of the entropy gives another that gives an estimate that is just as good. Similarly, conjugating an operator by a permutation doesn't change anything of note at all. So, there's nothing stopping you from using the average of all of these variants of the operator instead. Indeed, this can only decrease the uncertainty of the operator[12]. Nothing is lost by focusing on the invariant operators. But these are just sums over projectors into irreducible representations of the combined symmetry group.

Things simplify a bit more: the tensor product of n copies of $\mathbb{C}^{\otimes n}$ breaks up into

$$\mathbb{C}^{\otimes n} = \bigoplus_{\lambda \vdash n} S_\lambda \otimes V_\lambda. \tag{2.1}$$

This is a direct sum over λ, the partitions of n. These partitions label both the irreducible representations of S_n, S_λ and the irreducible representations of $U(N)$, V_λ[13]. Since each representation of the symmetric group appears exactly once, all symmetric operators are linear combinations of symmetric group projections. This turns out to be equivalent to symmetric linear combinations of the permutations. More on that later. Therefore the entropy must be obtained by studying how the tensor products of the density matrix distributes itself in the various representations of the symmetric group.

This seems easy enough. Ah, a light at the end of the tunnel. Unfortunately, this light is an oncoming train.

But first, a tour of the representation theory of the symmetric group.

[12]The argument comes down to the trace of a symmetric operator times an operator with no symmetric component being zero.

[13]When $N > n$, there are some partitions that do not give a representation of $U(N)$. This little fact is going to be swept under the rug to cause problems for the next residents of the property.

3 The Representation Theory of The Symmetric Group.

The following has been copied dutifully from my notes, which have been copied from Andrew Kobin's notes on representation theory. Perhaps, for a coherent and understandable account, you may want to look there.

Representations of finite groups are not too unlike representations of compact Lie groups. They are unitary and break up into a finite number of irreducible representations. One such representation is the regular representation: it consists of linear combinations of elements of the group. This is actually a representation of the product of two copies of the group: one acts on the left and one on the right. Given a representation V of the group, tensoring it with its dual V^*, that is, considering the linear maps from V to itself, gives another representation of the product of the groups. Each group element is associated with just such a map from V to itself so there is a canonical mapping from the regular representation $\mathbb{C}[G]$ to $V \otimes V^*$. A bit of careful logic ends up demonstrating that this actually decomposes the regular representation into a direct sum of squares of each of the irreducible representations.

$$\mathbb{C}[G] = \bigoplus_{\mu} V_\mu \otimes V_\mu^* \tag{3.1}$$

Given an irreducible representation λ and a conjugacy class of the group G, one can compute the trace $\mathrm{Tr}_{\lambda}(g)$ for any representative g of the conjugacy class. This is called the character $\chi_\lambda(g)$. The study of these characters is called *character theory*[14]. The product of traces is the trace on the tensor product. Let $|C_G(g)|$ be the the number of elements of G commuting with g. There are $|C_G(g)|$ different h's for which ghg'^{-1} to equals h if g and g' are in the same conjugacy class and no way for this to happen if they are not. Thus, some intense staring[15] will show that

$$\sum_{\lambda} \chi_\lambda(g)\overline{\chi_\lambda(g')} = \mathrm{Tr}_{\mathbb{C}[G]}(g \otimes g') = |C_G(g)|\mathbb{1}(g \sim g'). \tag{3.2}$$

Some similar fiddling around shows such an orthogonality relation when summing over g and so the characters $\chi_\lambda(g)$ can be rescaled to form a unitary matrix. This implies that the number of irreducible representations equals the number of conjugacy classes of the group and that the numbers of each irreducible representation in a representation is fixed entirely by the trace of each of the conjugacy classes.

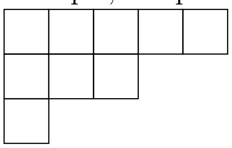
On to the symmetric groups. The number of conjugacy classes isn't too hard: they are classified by the partition of elements into cycles in the cycle decomposition. Since

[14] No relation to TvTropes or the Unicode consortium.

[15] This, along with similar phrases such as 'this is obvious', 'it is easy to see', and the perennial favourite 'trivial' are common proof techniques. These are especially popular among mathematics professors teaching students, as it is vital to teach the next generation of researchers the importance this technique holds.

the total number of elements permuted is n, this gives a one-to-one correspondence with the partitions of n[16]. The number of irreducible representations of the symmetric group must thus have to be the number of partitions of n. In fact, there is a canonical way to construct such an irreducible representation in terms of a partition.

Let $\lambda \vdash n$; λ is a partition of n, that is. Given λ one can construct a diagram where the number of boxes in each row is one of the components of the partition. For example, the partition $(5, 3, 1)$ gives the Ferrers diagram:



The boxes can be filled in with numbers 1 to n to give a Young tableau[17]:

2	6	5	8	9
3	1	4		
7				

A tabloid is given by treating different orderings of the elements in its rows as the same, so

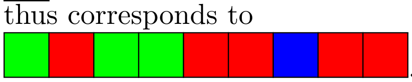
$$\begin{array}{ccccc} \hline 2 & 6 & 5 & 8 & 9 \\ \hline 3 & 1 & 4 & & \\ \hline 7 & & & & \\ \hline \end{array} = \begin{array}{ccccc} \hline 2 & 5 & 6 & 8 & 9 \\ \hline 1 & 3 & 4 & & \\ \hline 7 & & & & \\ \hline \end{array} = \begin{array}{ccccc} \hline 8 & 9 & 2 & 6 & 5 \\ \hline 4 & 3 & 1 & & \\ \hline 7 & & & & \\ \hline \end{array}. \tag{3.3}$$

Given a partition $\lambda \vdash n$, the representation of the permutation group given by linear combinations of tabloids of that shape gives M_λ, the tabloid representation of that shape[18]. It is important to note how the action of the permutation group permutes the *numbers*, not the *cells*. A different way to write down a tabloid is as a coloring of n cells: cell i is colored the color of row j if the number i is on row j in the tabloid.

The tabloid

2	5	6	8	9
1	3	4		
7				

thus corresponds to

.

The colorings with the desired number of each color are thus taken to each other under permutation of the cells.

[16]Unlike just about any other combinatorial thing of the form 'How many ways can you put n things in k boxes counting so and so as the same', there is no closed form for the partition functions. This is the first hint of things going wrong. The generating function for the partition function is up to some simple factors the Dedekind eta function and is the same as the number of states of a given level in a typical representation of the Virasoro algebra, counts the number of states in the free boson in two dimensions and other such deep connections. In other words, it's a royal pain.

[17]Plural: tableaux. Don't ask me why.

[18]This presumably keeps track of how various details of the personal lives of celebrities change under permutation.

This will be on the test later.

The number of tabloid representations, the number of partitions of n, is exactly the number of irreducible representations. However, they are not irreducible[19,20].

The actual irreducible representations are classified by *Specht modules*. The tabloids were almost right: The Specht modules live inside them. They are spanned by taking a given Young tableau, antisymmetrizing under permutations of the columns, and then converting this linear combination of Young tableaux into a linear combination of tabloids. Note that doing these operations in any other order is ill defined. Some long and lengthy shenanigans end up showing that this subspace is irreducible and gives a correspondence between irreducible representations and partitions.

Note how the symmetric combinations of the elements of a finite group correspond to the conjugacy classes of the group; there are the same number of these as projectors. The character table ensures that nothing silly happens, and so projectors can be expressed in terms of sums of group elements.

One last thing. There is an partial ordering on partitions called *dominance ordering*, with the partition with one component, corresponding to the trivial representation, on top and the partition consisting of all ones, corresponding to the antisymmetric representation[21] on bottom. The reducible tabloid representation splits up into irreducible representations, where each of the irreducible representations are above the tabloid representation on the dominance order. The multiplicity of the irreducible representation S_λ in the tabloid representation M_μ is the Kostka number $K_{\lambda\mu}$. This gives an triangular matrix with ones on the diagonals and is invertible.

Back to the matter at hand ...

4 On the Train Tracks

Density matrix ρ. n copies. Operator expectation $\mathrm{Tr}\left(\rho^{\otimes n}\mathcal{O}\right)$. Entropy what? Having now obsoleted the preceding sections, let us continue on.

A density matrix can be diagonalized, so it's easy to put the density matrix in the form of a diagonal matrix with probability p_i on state i[22]. The space of states is spanned by configurations of the form $|k_1\rangle \otimes |k_2\rangle \otimes \ldots \otimes |k_n\rangle, k_i \in 1, \ldots, N$. There are no off diagonal elements of the density matrix in this basis and this actually reduces to a classical probability question. Quantum effects are in retreat, we can advance with

[19] A simple demonstration of this fact can be seen by noting that the rest of the paper would have resolved in a beautiful, elegant, and conceptually satisfying manner were this to be true. Therefore, Murphy's theorem proves it false.

[20] An *actual* demonstration of this fact comes from the one dimensional trivial representation being found in each of the tabloid representations, which are not generically one dimensional.

[21] Would this be the anti-trivial representation?

[22] Except, of course, when there is any reason to be interested in the entropy in the first place.

confidence, our system will be fully analyzed in short order, a chicken in every pot, and world peace is just around the corner. Of course.

The probability for being in the state $|k_1, \dots, k_n\rangle$ is $p_{k_1} \dots p_{k_n}$. These states get shuffled into each other under permutations. Subsets of this set of states with a given number of the copies in a given state are taken to each other under permutation. This is exactly a tabloid representation with the partition given by the state counts sorted by number. So the state $|1, 3, 2, 4, 1, 3, 3\rangle$ is in the sector with two copies in state 1, one in state 2, three in state 3 and one in state 4. The sector is isomorphic to the tabloid representation with partition $3, 2, 1, 1$. The probability that there are n_i states in state i is thus

$$\frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}. \tag{4.1}$$

The two statistics that are familiar to us all are the Bose statistics and Fermi statistics. Fermions, due to the Pauli exclusion principle, refuse to hanky-panky with each other by occupying the same states as each other. Bosons, however, are rather … gregarious. There are anyons and other such things too, but we will not be treading into that little bear trap today[23]. The occupancy statistics of a bosonic and a fermionic state follow the distribution of a Bernoulli trial and a geometric distribution, irrespectively[24]. These are both quantum statistics, so what's the classical version? It's the Poisson distribution, with probability mass function

$$P(n; \lambda) = e^{-\lambda} \frac{\lambda^n}{n!}, \; n \in \mathbb{N}. \tag{4.2}$$

One annoyance that has not been dealt with so far is the fact that all the density matrices are normalized, which is rather inconvenient when the the normalizing factor contains all the information that one may be trying to figure out. Luckily, the Poisson distribution plays nicely. Consider an unnormalized density matrix $\tilde{\rho}$. The normalized one is $\rho = \tilde{\rho}/\mathrm{Tr}\,(\tilde{\rho})$. So far n has been held constant. But now let n be distributed according to a Poisson distribution with parameter $\mathrm{Tr}\,(\tilde{\rho})\lambda$, giving a probability distribution going as

$$e^{-\mathrm{Tr}\,(\tilde{\rho})\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n \mathrm{Tr}\,(\tilde{\rho})^n}{n!} \rho^{\otimes n} = e^{-\mathrm{Tr}\,(\tilde{\rho})\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \tilde{\rho}^{\otimes n} \tag{4.3}$$

The partition functions denominators have nicely canceled, while taking an expectation value for large λ averages over the expectation values with the normalized ρ for large n. The probability of the state having n_i copies in state i is

$$e^{-\mathrm{Tr}\,(\tilde{\rho})\lambda} \frac{\lambda^n}{n!} n! \prod_i \frac{\tilde{p}_i^{n_i}}{n_i!} = \prod_i e^{-\tilde{p}_i \lambda} \frac{\lambda^n \tilde{p}_i^n}{n_i!}. \tag{4.4}$$

[23] I have the premonition that I might have to thoroughly confuse myself with some generalized version of anyons someday.

[24] Giving two lists that needs to be matched and then tacking 'respectively' at the end has often caused no end of confusion. Therefore, not doing that must not be confusing, right?

It factorizes into a separate little Poisson for each state! A situation where this kind of Poisson might occur is in the following. Considering adding an additional state $|e\rangle$ to the Hilbert space, bringing it to $\mathbb{C}\oplus\mathbb{C}^N$. Then one possible density matrix to construct is

$$\left(1+\frac{\lambda \mathrm{Tr}\,\tilde{\rho}}{M}\right)^{-1}\left(|e\rangle\langle e|+\frac{\lambda}{M}\tilde{\rho}\right) \tag{4.5}$$

for some *very* large integer M. Then, consider the tensor product of M copies of that state. Measuring whether a state is in $|e\rangle$ or not commutes with everything of interest, so the previous operators can be translated to here by applying them to a state with a pre-specified configuration of e's and then summing over the projectors onto those configurations. By construction, the number of not-e's is distributed exactly corresponding to a Poisson distribution, so it reduces to the earlier case. See also how the number of copies in each specific state naturally sorts into independent Poissons with the right rates. Amusingly enough, the von Neumann entropy of this state comes out to something sensible in terms of the von Neumann entropy of the original density matrix while the dependence on the Renyi entropy goes to zero as M is taken to infinity. Another example of power counting and the replica trick doing fishy things in back alleys.

The Poisson process above gives the occupancy counts of each state. The possible assignments with a given occupancy count are taken to each other by permutation and so are in some sense irreducible *sets*. If this was a bucket of dice, occupancy counts would be exactly the correct kind of thing to look at. However, these are density matrices. Lumping all terms that give the same tabloid representation already gives some nasty correlations and even that is not the end of it. What the measurement actually separates is the irreducible representations, which are even more of a pain to deal with. This makes calculations of such things as the uncertainties and suchlike a right pain.

What *can* be done is to calculate the physicist's old friend, expectation values of operators. The whole start of the mess came from the Renyi estimates, so let's calculate those. The trace of a cyclic shift is the sum over colorings that the shift leaves invariant. In the cycle, all the colors have to be the same. The number of ways that this can happen is

$$\sum_i \frac{(n-k)!}{(n_i-k)!\prod_{j\neq i} n_j!}. \tag{4.6}$$

This gives the expectation value

$$\sum_i \frac{n_i!(n-k)!}{(n_i-k)!n!}, \tag{4.7}$$

where both of the sums over i are restricted to where $n_i \geq k$. Note how this approaches the dice bucket counting for the exponential of the Renyi entropy when n_i is large. By construction, the average value of this estimate gives the correct value of $\mathrm{Tr}\left(\rho^k\right)$ when $n \geq k$. The $n < k$ case has to be dropped for sensibility's sake. As λ becomes large, the probability of being in a state where $n < k$ drops about exponentially, so this gives the average over the Poisson case as approaching the correct value exponential-ish in $\mathrm{Tr}\left(\tilde{\rho}\right)\lambda$. Now, the von Neumann entropy can be found at the derivative when $n = 1$, so there is nothing preventing *considering* the value of the estimate there.

Some good ol' fashioned bashing with the properties of the gamma and digamma functions gives the formula for the entropy estimate coming from analytic continuation as

$$\sum_i \sum_{m=n_i}^{n-1} \frac{1}{m}\frac{n_i}{n}. \tag{4.8}$$

My old nemesis for the duration of this jaunt, the replica trick, has subtly slipped in an attempt at sabotage. There is an ambiguity when n_i is zero, where one has to decide what happens when m hits zero. The analytic continuation specifies that the n_i/m should go to one. This improves the asymptotics of the entropy estimate *when the dimension of the Hilbert space is finite.* In a infinite dimensional space, or at least the case when there are a significant fraction of the states unfilled even when $\lambda\mathrm{Tr}\left(\tilde{\rho}\right)$ is large, this term gives the *wrong* estimate for the entropy. When using sketchy tricks, let the buyer beware.

So let's do our due diligence and show that the corrected entropy estimate converges to the right answer. The average over the Poisson distribution is such that the asymptotics in n and the asymptotics in $\lambda\mathrm{Tr}\left(\tilde{\rho}\right)$ can be recovered from each other. And the Poisson averaged case is so much easier to calculate.

A single term of the estimate is

$$\sum_{m=n_i}^{(n-n_i)+n_i-1} \frac{1}{m}\frac{n_i}{(n-n_i)+n_i} \tag{4.9}$$

Crucially, it depends only on n_i and $n_\perp = n - n_i$. But the distribution of these two variables is known: n_i and $n_\perp$ follow two independent Poisson distributions. The rate for n_i is $\lambda\widetilde{p_i}$ and the rate for $n_\perp$ is $\lambda\widetilde{p_\perp}$, where $\widetilde{p_\perp} = \mathrm{Tr}\left(\widetilde{\rho}\right) - \widetilde{p_i}$. Instead of doing N sums, we only need two.

One may be tempted to leave the remaining calculation as an exercise to the reader, but I won't do that here out of the goodness of my heart[25]. The averaged estimate is

$$\sum_{n_i=0}^{\infty}\sum_{n_\perp=0}^{\infty} e^{-\lambda(\widetilde{p_i}+\widetilde{p_\perp})}\frac{(\lambda\widetilde{p_i})^{n_i}(\lambda\widetilde{p_\perp})^{n_\perp}}{n_i!n_\perp!}\sum_{m=n_i}^{n_i+n_\perp-1}\frac{1}{m}\frac{n_i}{n_\perp+n_i}. \tag{4.10}$$

Reshuffling the indices with $n_i = u$, $m - n_i = v$, $n - m = w$ gives

$$e^{-\lambda\mathrm{Tr}\left(\tilde{\rho}\right)}\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\sum_{w=0}^{\infty}\lambda^{u+v+w}\widetilde{p_i}^{u}\widetilde{p_\perp}^{v+w}\frac{1}{u!(v+w)!}(1-\delta_w)\frac{u}{u+v}\frac{1}{u+v+w}. \tag{4.11}$$

[25] I need to refer to an intermediate step in the calculation later, that is.

This is the kind of expression that can be solved by whacking it over the head with generating functions until it confesses. The last three sub-terms on the right turn into a rather fiddly two parameter integral over $a^u b^v c^w$. If one denotes $f(a,b,c) = a^u b^v c^w$, then

$$\begin{aligned}
&(1-\delta_w)\frac{u}{u+v}\frac{1}{u+v+w} && (4.12)\\
&\qquad = \int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \; (1-\delta_w)\, u\, x^{u+v-1} y^{u+v+w-1} && (4.13)\\
&\qquad = \int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y \; \partial_a (f(xy,xy,y) - f(xy,xy,0)). && (4.14)
\end{aligned}$$

Note how this is using the $0/(0+0) = 0$ version of the estimate.

The generating function is

$$e^{-\lambda \mathrm{Tr}\,(\tilde{\rho})} e^{\lambda \widetilde{p_i} a} \frac{1}{b-c}\left(e^{\lambda \widetilde{p_\perp} b} b - e^{\lambda \widetilde{p_\perp} c} c\right). \tag{4.15}$$

The rather peculiar form of the expression comes from the factorials in the denominator has a $(v+w)!$ instead of a $v!w!$, giving a truncated geometric series in the sum when $v+w$ is held fixed. Note that without this, there is no way for logs to come up.

Now, put all the ingredients into a pot then heat to a low simmer.

$$\begin{aligned}
&\int_0^1 \int_0^1 \mathrm{d}x\,\mathrm{d}y\, \lambda \widetilde{p_i} e^{-\lambda \mathrm{Tr}\,(\tilde{\rho})} e^{\lambda \widetilde{p_i} xy} \\
&\qquad \left[\frac{1}{y-xy}\left(e^{\lambda \widetilde{p_\perp} y} y - e^{\lambda \widetilde{p_\perp} xy} xy\right) - \frac{1}{xy}\left(e^{\lambda \widetilde{p_\perp} xy} xy - 0\right)\right] && (4.16)\\
&= \lambda \widetilde{p_i} \int_0^1 \int_0^1 \mathrm{d}x\,\mathrm{d}y\, \frac{1}{1-x} e^{-\lambda(1-y)\mathrm{Tr}\,\tilde{\rho}} \left(e^{-y(1-x)\widetilde{p_i}} - e^{-y(1-x)\mathrm{Tr}\,\tilde{\rho}}\right) && (4.17)
\end{aligned}$$

Stir for 10-20 minutes until curdling occurs.

$$= \lambda \widetilde{p_i} \int_0^1 \int_0^1 \mathrm{d}t\,\mathrm{d}s \frac{1}{t} e^{-\lambda s \mathrm{Tr}\,\tilde{\rho}} (e^{-(1-s)t\widetilde{p_i}} - e^{-(1-s)t\mathrm{Tr}\,\tilde{\rho}}) \tag{4.18}$$

Drain the curds, then press into cheese.

$$\begin{aligned}
&= \frac{\widetilde{p_i}}{\mathrm{Tr}\,\tilde{\rho}} \int_0^1 \int_0^1 \mathrm{d}s\,\mathrm{d}t e^{-\lambda s \mathrm{Tr}\,\tilde{\rho}} (1-s) \int_{\frac{\widetilde{p_i}}{\mathrm{Tr}\,\tilde{\rho}}}^1 e^{-(1-s)tw}\,\mathrm{d}w && (4.19)\\
&= \frac{\widetilde{p_i}}{\mathrm{Tr}\,\tilde{\rho}} \int_{\frac{\widetilde{p_i}}{\mathrm{Tr}\,\tilde{\rho}}}^1 \left[\frac{1}{w}\left(1 - e^{-\lambda \mathrm{Tr}\,\tilde{\rho} w}\right) - \frac{1}{1-w}\left(e^{-\lambda \mathrm{Tr}\,\tilde{\rho} w} - e^{-\lambda \mathrm{Tr}\,\tilde{\rho}}\right)\right] \mathrm{d}w && (4.20)
\end{aligned}$$

Store in a cool, dry place. Choking hazard, small parts. Not suitable for children under the age of three.

Anyways ...

The first of the four terms gives the exact value for that component of the entropy.

The rest give corrections. The $1/(1-w)$ terms give a correction of order $\frac{\widetilde{p}_i}{\mathrm{Tr}\,\tilde{\rho}}\frac{1}{\lambda\mathrm{Tr}\,\tilde{\rho}}e^{-\lambda\widetilde{p}_i}$, which amounts to asymptotically small potatoes in any limit[26]. The $w^{-1}\exp(-\lambda\mathrm{Tr}\,\tilde{\rho}w)$ term is more interesting. It works to cancel the 1 at small w's, causing the log term to, instead of soaring high and free like it wants, level off when $\lambda\widetilde{p}_i$ starts to drop below one and n_i starts having a significant probability of being zero. This acts as an effective cutoff on the probabilities contributing to the entropy. See how the effect of this drops out as $\lambda\mathrm{Tr}\,\tilde{\rho}$ increases because of the convergence of the entropy. Asymptotically, the term limits to $e^{-\lambda\widetilde{p}_i}$, which might seem like a problem since the $\widetilde{p}_i/\mathrm{Tr}\,\tilde{\rho}$ has canceled, but the formula is only valid for $\lambda\widetilde{p}_i$ large. Indeed this 'problem' is exactly what the extra term that the replica trick was trying to sneak in was there to cancel.

There is the question, however, of what bloody operator the damn estimate is supposed to correspond to. This is not so bad. Fixing a value of n_i and n gives a contribution, when the $\frac{1}{n!}(\lambda\mathrm{Tr}\,\tilde{\rho})^n$ is taken out, of $p_i^{n_i}(1-p_i)^{n-n_i}$. Summing over i makes this $\mathrm{Tr}\,(\rho^{n_i}(1-\rho)^{n-n_i})$, making the relevant operator the sum over conjugation of the appropriate cycle permutations. One might then ask further questions, like what the uncertainty of the estimate is, how the estimate, which has the same binomial coefficient problem as earlier, might be calculated from the path integral, and how the value breaks up over the actual, measured, irreducible representations[27]. At this point, everything nice that has just happened just stops working and a cloud of darkness falls upon the whole enterprise.

The most basic question is what the probabilities are for finding $\rho^{\otimes n}$ in a given representation. This doable, but already annoyingly nontrivial. The probability of landing in a given tabloid representation is the sum of permutations of $p_1^{n_1}\dots p_k^{n_k}$ times the dimension of the representation. A symmetric polynomial, that is. The Kostka numbers which dictate the way tabloids break up into irreducibles can be written in terms of such polynomials and some shuffling gives the expression for the probability as

$$P_\lambda = \dim_\lambda s_\lambda(p_1, p_2, \dots). \tag{4.21}$$

s_λ is a symmetric polynomial known as the *Schur polynomial*, which is a ratio of two determinants[28]. The denominator is the measure that comes up in random matrix theory, oddly enough.

What does the entropy estimate look like on the individual irreducible representations? I don't know. What's the uncertainty of the entropy estimate? Well it's related to the average of the squares of the coefficients on the irreducibles. Perhaps it could be analyzed by multiplying out the conjugacy classes by hand, but that turns into a pretty quagmire after a bit when the cycles get big. So, I don't know. One hopes to do QES type things or condensed matter physics or suchlike using this. For quantum

[26]These potatoes are especially prized in haute cuisine, vanishing in the limit as food approaches the table.

[27]I'm hoping to reach eight commas in a sentence someday.

[28]At least when $N = n$. I'm not sure. Schur. Ha.

gravity calculations, this requires knowing the physical properties of such things as how a defect line attached to permutation group projectors acts like in the replica trick when there is a *very* large number of replicas. How could you do that? I don't know.

For that matter, what does the probability distribution over partitions even look like at large n? I don't know. However, it looks like someone else does. There seems to be a paper about the asymptotics of the Schur polynomial[29]. Highlights include random matrix theory, the GUE, an infinite dimensional unitary group, q-deformations thereof, and a random lozenge. A paper on the related topic of the character table of the symmetric group at large n[30] offers non-crossing partitions, non-commutative probability, a genus expansion over two dimensional surfaces, and the Jucys-Murphy element. This would all be very useful if I actually understood any of this.

At this point, one may be tempted go give up, so that is exactly what I did. Good night and I hope you enjoyed the show.

[29] arXiv: 1301.0634
[30] arXiv: math/0304275

5 Appendix A

N