Expansion of Sum over Disjoint Indices

Chanwoo Chun

April 2, 2025

1 Problem statement

Many estimators can be written in the following form:

$$V_n \coloneqq \frac{1}{\text{number of summands}} \sum_{i_1 \neq i_2 \neq \ldots \neq i_n} T_{i_1, i_2, \ldots, i_n},$$

where T is a rank-n tensor. In this form, it is difficult to compute this sum in practice. The question is: how can we expand V into a sum of "regular" sums, as opposed to sums over disjoint indices? For example:

$$\sum_{i \neq j} K_{ij} = \sum_{ij} K_{ij} - \sum_{i} K_{ii}.$$

In higher rank tensor ("rank" here means the number of indices), the coefficients of the expansion become highly non-trivial.

One motivating example is the second spectral moment estimator of a kernel integral operator, which is briefly explained in the appendix.

1.1 Starting framework with an example

Denote by Z the number of summands and n the rank of the tensor, i.e., the number of indices. We can rewrite V as

$$V_n \coloneqq \frac{1}{Z} \sum_{i_1, i_2, \ldots, i_n} T_{i_1, i_2, \ldots, i_n} D_{i_1, i_2, \ldots, i_n}$$

where

$$D_{i_1, i_2, \ldots, i_n} \coloneqq \prod_{1 \leq a < b \leq n} \left(1 - \delta_{i_a i_b}\right)$$

The role of $D_{i_1, i_2, \ldots, i_n}$ is to be 0 when at least one index is connected. For example:

$$V_3 \coloneqq \frac{1}{Z} \sum_{i_1, i_2, i_3} T_{i_1, i_2, i_3} \left(1 - \delta_{i_1 i_2}\right) \left(1 - \delta_{i_1 i_3}\right) \left(1 - \delta_{i_2 i_3}\right)$$

Next, if we expand the polynomial entry of D, it becomes the instruction of how to contract the indices. Continuing on the V_3 example, we get

$$D_{i_1, i_2, i_3} = 1 - \delta_{i_1 i_2} - \delta_{i_2 i_3} - \delta_{i_1 i_3} + \delta_{i_2 i_3}\delta_{i_1 i_2} + \delta_{i_1 i_3}\delta_{i_1 i_2} + \delta_{i_1 i_3}\delta_{i_2 i_3} - \delta_{i_1 i_3}\delta_{i_2 i_3}\delta_{i_1 i_2} \tag{1}$$

Here, it will be essential to view each term as a graph. In this graph, each node/vertex represents an index, and each edge represents a contraction $\delta_{i_a i_b}$ with the weight -1. From this graph, it is obvious that the last four terms are all "equivalent", i.e. contract the same set of indices:

$$\delta_{i_1 i_2 i_3} \coloneqq \delta_{i_2 i_3}\delta_{i_1 i_2} \equiv \delta_{i_1 i_3}\delta_{i_1 i_2} \equiv \delta_{i_1 i_3}\delta_{i_2 i_3} \equiv \delta_{i_1 i_3}\delta_{i_2 i_3}\delta_{i_1 i_2}$$

which we denote as $\delta_{i_1 i_2 i_3}$ which is 1 iff all indices in the subscript match and 0 otherwise. However, the sum of the edge weights can be different depending on the parity of the number of edges. Taking these into account, we get the following simplification:

$$D_{i_1, i_2, i_3} = 1 - \delta_{i_1 i_2} - \delta_{i_2 i_3} - \delta_{i_1 i_3} + 2\delta_{i_1 i_2 i_3} \tag{2}$$

Now each contraction in 2 defines a unique summation term in the expanded form of V_3:

$$V_3 = \frac{1}{Z}\left(\sum_{i_1,i_2,i_3} T_{i_1,i_2,i_3} - \sum_{i_1,i_2} T_{i_1,i_1,i_2} - \sum_{i_1,i_2} T_{i_1,i_2,i_2} - \sum_{i_1,i_2} T_{i_1,i_2,i_1} + 2\sum_{i} T_{i,i,i}\right),$$

which can be further simplified if we assume T is symmetric in some sense, e.g. $T_{i_1,i_1,i_2} = T_{i_1,i_2,i_2} = T_{i_1,i_2,i_1}$, but for now we leave it like above to keep the framework general. Now the natural question is, how does 2 generalize to a higher rank tensor? What is the coefficient for each contraction in 2?

2 The solution

2.1 Counting the number of regular summations

First, let us compute how many possible contractions there are for n indices. This is equivalent to the number of partitions of a set of size n (with labeled elements), which is given by the Bell number B_n. Therefore B_n is the total number of regular summations ($\sum$'s) we will get in our sum expansion. Note that B_n grows exponentially. Now let us figure out the coefficient for each regular summation, i.e. each contraction.

2.2 (Main) Getting the coefficient for each regular summation

Counting the number of connected graphs

In general, a single partition of the indices, e.g. $\delta_{i_1,i_4}\delta_{i_6 i_8 i_9}$, can be represented by a graph with multiple connected subgraphs, e.g. δ_{i_1,i_4} and $\delta_{i_6 i_8 i_9}$ are connected subgraphs. In this section, let us first focus on one connected graph, i.e. one connected contraction. Denote by $\mathcal{G} := \{G_i\}$ the set of all connected graphs over a vertex set V, with each graph denoted by $G_i = (V, E_i)$ where E_i is the set of edges. The following is the coefficient for a single connected contraction:

$$\sum_{i=1}^{|\mathcal{G}|} (-1)^{|E_i|}.$$

After highly non-trivial calculations, we can get a closed-form solution:

$$\sum_{i=1}^{|\mathcal{G}|} (-1)^{|E_i|} = (-1)^{k-1}(k-1)!$$

where $k := |V|$ is the number of vertices. The derivation uses an exponential generating function and "exponential formula" from statistical mechanics and (apparently) quantum field theory. The full derivation is in the appendix.

Accounting for multiple connected graphs

Any graph is a collection of connected subgraphs. The coefficient for the summation/overall contraction corresponding to multiple connected subgraphs is then given by:

$$\prod_{i=1}^{\#\text{ of connected subgraphs}} (-1)^{k_i-1}(k_i-1)!$$

where k_i is the number of vertices in the ith subgraph.

Proof for this is simple, and is a part of the previous proof, so it can be found in the appendix. However, we repeat it here for clarity. Say $\mathcal{G}$ is a set of graphs that is composed of h connected graphs over vertex sets $V^{(1)}, V^{(2)}, \ldots, V^{(h)}$. Let $\mathcal{G}^{(j)}$ be a set of all connected graphs over $V^{(j)}$ and $|E_i^{(j)}|$ be the number of edges in ith graph in $\mathcal{G}^{(j)}$. Then the coefficient should be:

$$\sum_{i=1}^{|\mathcal{G}|} (-1)^{\sum_{l=1}^{h}\left|E_i^{(l)}\right|}$$

which can be decomposed as

$$\sum_{i=1}^{|\mathcal{G}^{(1)}|}\sum_{j=1}^{|\mathcal{G}^{(2)}|}\cdots\sum_{k=1}^{|\mathcal{G}^{(h)}|} (-1)^{\left|E_i^{(1)}\right|}(-1)^{\left|E_j^{(2)}\right|}\cdots(-1)^{\left|E_k^{(h)}\right|},$$

which is equivalent to

$$\left(\sum_{i=1}^{|\mathcal{G}^{(1)}|}(-1)^{\left|E_i^{(1)}\right|}\right)\left(\sum_{i=1}^{|\mathcal{G}^{(2)}|}(-1)^{\left|E_i^{(2)}\right|}\right)\left(\sum_{i=1}^{|\mathcal{G}^{(3)}|}(-1)^{\left|E_i^{(3)}\right|}\right)\cdots,$$

which means

$$\sum_{i=1}^{|\mathcal{G}|}(-1)^{\sum_{l=1}^{h}\left|E_i^{(l)}\right|}=\prod_{i=1}^{h}(-1)^{|V^{(i)}|-1}\left(|V^{(i)}|-1\right)!.$$

Q.E.D.

Final expansion formula

In conclusion, we get the following expansion as the solution to our problem:

$$D_{i_1,i_2,\ldots,i_n}\coloneqq\prod_{1\le a<b\le n}(1-\delta_{i_a i_b})$$

$$=\sum_{h=1}^{n}\sum_{\left\{V^{(1)},\ldots V^{(h)}\right\}}\left(\prod_{i=1}^{h}(-1)^{|V^{(i)}|-1}\left(|V^{(i)}|-1\right)!\right)\left(\prod_{i=1}^{h}\delta_{V^{(i)}}\right)$$

where $\sum_{h=1}^{n}\sum_{\left\{V^{(1)},\ldots,V^{(h)}\right\}}$ simply means sum over all partitions of the indices. Here, $V^{(i)}$ is one block of a partition, $\left\{V^{(1)},\ldots,V^{(h)}\right\}$ is one partition, and $\sum_{\left\{V^{(1)},\ldots,V^{(h)}\right\}}$ means sum over all partitions with h number of blocks.

3 Generalization to multiple disjoint sets of indices

The problem that we will deal with requires further generalization. In one tensor, there can be multiple groups of indices that need to be disconnected/disjoint within their own respective groups:

$$V_n\coloneqq\frac{1}{Z}\sum_{\left(i_1^{(1)}\neq\ldots\neq i_{n^{(1)}}^{(1)}\right)}\sum_{\left(i_1^{(2)}\neq\ldots\neq i_{n^{(2)}}^{(2)}\right)}\cdots\sum_{\left(i_1^{(p)}\neq\ldots\neq i_{n^{(2)}}^{(p)}\right)}T_{i_1^{(1)},\ldots,i_{n^{(p)}}^{(p)}},$$

This is motivated by the following example. Consider a matrix $X\in\mathbb{R}^{P\times Q}$. The following sum is the unbiased estimate of the second spectral moment of an infinitely large matrix from which X is randomly sampled by sampling the rows and columns independently:

$$\frac{1}{P\left(P-1\right)Q\left(Q-1\right)}\sum_{i\neq j}\sum_{\alpha\neq\beta}X_{i\alpha}X_{j\alpha}X_{i\beta}X_{j\beta}.$$

Using the above notation, $i_1^{(1)}=i$, $i_2^{(1)}=j$, $i_1^{(2)}=\alpha$, $i_2^{(2)}=\beta$ and $T_{i_1^{(1)},i_2^{(1)},i_1^{(2)},i_2^{(2)}}=X_{i_1^{(1)}i_1^{(2)}}X_{i_2^{(1)}i_1^{(2)}}X_{i_1^{(1)}i_2^{(2)}}X_{i_2^{(1)}i_2^{(2)}}$, we have:

$$\frac{1}{P\left(P-1\right)Q\left(Q-1\right)}\sum_{i_1^{(1)}\neq i_2^{(1)}}\sum_{i_1^{(2)}\neq i_2^{(2)}}T_{i_1^{(1)},i_2^{(1)},i_1^{(2)},i_2^{(2)}}.$$

Note that

$$V_n\coloneqq\frac{1}{Z}\sum_{\left(i_1^{(1)}\neq\ldots\neq i_{n^{(1)}}^{(1)}\right)}\sum_{\left(i_1^{(2)}\neq\ldots\neq i_{n^{(2)}}^{(2)}\right)}\cdots\sum_{\left(i_1^{(p)}\neq\ldots\neq i_{n^{(2)}}^{(p)}\right)}T_{i_1^{(1)},\ldots,i_{n^{(p)}}^{(p)}}\prod_{r=1}^{p}D_{i_1^{(r)},\ldots,i_{n^{(r)}}^{(r)}}$$

Therefore,

$$\prod_{r=1}^{p}D_{i_1^{(r)},\ldots,i_{n^{(r)}}^{(r)}}\coloneqq$$

$$=\prod_{r=1}^{p}\sum_{h=1}^{n^{(r)}}\sum_{\left\{V^{(1)},\ldots V^{(h)}\right\}}\left(\prod_{i=1}^{h}(-1)^{|V^{(i)}|-1}\left(|V^{(i)}|-1\right)!\right)\left(\prod_{i=1}^{h}\delta_{V^{(i)}}\right)$$

the notations here assume that $\left\{V^{(1)},\ldots V^{(h)}\right\}$ is a partition of a disjoint index set that is specified by r in the outermost product $\prod_{r=1}^{p}$.

4 The algorithm

The implementation is now straight-forward.

5 Appendix

5.1 Motivating example

For example, the square integral of function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ w.r.t. some probability measure $\rho_{\mathcal{X}}$ over $\mathcal{X}$, i.e. $\left\langle k(x,y)^2 \right\rangle_{x,y\sim\rho_{\mathcal{X}}}$ (assume this is finite), can be estimated from P number of finite samples of $x \sim \rho_{\mathcal{X}}$ without bias with the following sum:

$$\frac{1}{P\,(P-1)} \sum_{i_1 \neq i_2} k(x_{i_1}, x_{i_2})^2,$$

in which case T is a rank-2 tensor, i.e. a matrix, whose ijth entry is given by $k(x_i, x_j)$. To provide some context, $\left\langle k(x,y)^2 \right\rangle_{x,y\sim\rho_{\mathcal{X}}}$ is the sum of the squares of the eigenvalues of the kernel integral operator M_k implicitly defined as $M_k f = \int d\rho_{\mathcal{X}}(x)\, k(\cdot, x) f(x)$. It is immediately obvious that the above sum can be expanded like

$$\frac{1}{P\,(P-1)} \left(\sum_{i_1, i_2} k(x_{i_1}, x_{i_2})^2 - \sum_{i} k(x_i, x_i)^2 \right).$$

5.2 Proof using the exponential generating function

Denote by $\mathcal{G} \coloneqq \{G_i\}$ the set of all connected graphs over a vertex set V, with each graph denoted by $G_i = (V, E_i)$ where E_i is the set of edges. Here we prove

$$\sum_{i=1}^{|\mathcal{G}|} (-1)^{|E_i|} = (-1)^{k-1}\,(k-1)!$$

We will use the "exponential formula" from statistical mechanics. I got a hint from the question in Mathematics Stack Exchange.

Consider a series:

$$f(x) = a_1 x + \frac{a_2}{2} x^2 + \frac{a_3}{6} x^3 + \cdots + \frac{a_n}{n!} x^n + \cdots$$

whose coefficients a_k corresponds to the quantity what we want to find, which is [the sum of [the products of the edge weights in a given connected graph] over all possible connected graphs with k number of vertices], i.e. $\sum_{i=1}^{|\mathcal{G}|} (-1)^{|E_i|}$ where $|V| = k$.

According to the exponential formula, the exponential of $f(x)$ has a very interesting series:

$$g(x) \coloneqq e^{f(x)} = \sum_{n=0}^{\infty} \frac{b_n}{n!} x^n,$$

$$b_n \coloneqq \sum_{k} \sum_{\{V_1,\ldots,V_k\}} a_{|V_1|} \cdots a_{|V_k|}$$

where $\{V_1, \ldots, V_k\}$ is one partition of the set $\{1, \ldots, n\}$, and S_i is the ith block of the set. Therefore, the summation $\sum_k \sum_{\{V_1,\ldots,V_k\}}$ means sum over all partitions of the set $\{1, \ldots, n\}$. Let us first understand what the summand $a_{|V_1|} \cdots a_{|V_k|}$ signifies in our problem. It computes the following: first partition a graph with k blocks of vertices. There can be multiple connected graphs that connect the vertices in each block. The immediate interpretations of $a_{|V_1|} \cdots a_{|V_k|}$ is then, it is [the product of [the sum of [the products of the edge weights in a given connected graph] over all possible connected graphs that connected all vertices in a given block] over all k blocks]. This turns out to be equivalent to [the sum of [the products of the edge weights in the entirety given graph] over all possible graphs of n vertices with specific k blocks of connected vertices]. The following is a short proof of this statement: Say $\mathcal{G}$ is a set of graphs that is composed of k connected graphs over a given disjoint vertex sets $V^{(1)}, V^{(2)}, \ldots, V^{(k)}$. Let $\mathcal{G}^{(j)}$ be a set of all connected graphs over $V^{(j)}$ and $|E_i^{(j)}|$ be the number of edges in ith graph instance in $\mathcal{G}^{(j)}$. Then the $a_{|V_1|} \cdots a_{|V_k|}$ is

$$\left(\sum_{i=1}^{|\mathcal{G}^{(1)}|} (-1)^{\left|E_i^{(1)}\right|} \right) \left(\sum_{i=1}^{|\mathcal{G}^{(2)}|} (-1)^{\left|E_i^{(2)}\right|} \right) \cdots \left(\sum_{i=1}^{|\mathcal{G}^{(k)}|} (-1)^{\left|E_i^{(k)}\right|} \right)$$

$$= \sum_{\alpha=1}^{|\mathcal{G}^{(1)}|} \sum_{\beta=1}^{|\mathcal{G}^{(2)}|} \cdots \sum_{\gamma=1}^{|\mathcal{G}^{(k)}|} (-1)^{\left|E_\alpha^{(1)}\right|} (-1)^{\left|E_\beta^{(2)}\right|} \cdots (-1)^{\left|E_\gamma^{(k)}\right|},$$

$$= \sum_{\alpha=1}^{|\mathcal{G}|} (-1)^{\sum_{i=1}^{k} \left|E_\alpha^{(i)}\right|}$$

Now let us understand what $b_n \coloneqq \sum_k \sum_{\{V_1,\ldots,V_k\}} a_{|V_1|} \cdots a_{|V_k|}$ means. The following fact is useful here: any graph is a collection of connected subgraphs. From this we now know that b_n is simply [the sum of [the products of the edge weights in a given graph] over all graphs with n vertices]. b_n is much simpler to compute, since it sums over not just the connected graphs, but all possible graphs. It is easy to see that for $n \geq 2$,

$$b_n = \sum_{r=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{r} (-1)^r$$

Using the following relationship: $(x+y)^n \equiv \sum_{k=0}^{n} \binom{n}{k} x^k y^{(n-k)}$, we get: $b_n = (-1+1)^{\binom{n}{2}} = 0$ for $n \geq 2$. When $n = 1$, there is one vertex, and no edge. It can be seen that we have been treating "no edge" with weight 1, so $b_1 = 1$. The same is true when $n = 0$, i.e. $b_0 = 1$. Therefore,

$$g(x) = 1 + x$$

Now we can use the trick

$$f(x) = \log g(x)$$

which means

$$f(x) = \log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$$

Earlier we defined $f(x)$ as

$$f(x) = \sum_{n=1}^{\infty} \frac{a_n}{n!} x^n$$

Demanding both series to be equal, we get

$$\frac{a_n}{n!} = \frac{(-1)^{n-1}}{n}$$

$$a_n = (-1)^{n-1} (n-1)!$$

Therefore, [the sum of [the products of the edge weights in a given connected graph] over all possible connected graphs with n number of vertices] is $(-1)^{n-1} (n-1)!$. Q.E.D.

Apparently, there are many ways to arrive at this solution using Möbius function (link), or Principle of Inclusion-Exclusion, or cluster expansion in statistical mechanics (link).