Combinatorics - Lecture 1
Basics

Dr. Chintan Kr Mandal

Combinatorics leads to *Combinatorial Algorithms* which is the study of Combinatorial Structures. These can be informally classified as

- **Generation :** A generation algorithm will list all the objects under consideration in a certain order, such as a lexicographic order. It may be desirable to predetermine the position of a given object in the generated list without generating the whole list. Examples of the combinatorial structures we might wish to generate include subsets, permutations, partitions, trees and Catalan families.

- **Enumeration :** *Compute the number of different structures o f a particular type.* Every generation algorithm is also an enumeration algorithm, since each object can be counted as it is generated. The converse is not true, however. It is often easier to enumerate the number of combinatorial structures of a particular type than it is to actually list them. For example, the number of k-subsets of an n-element set is

$$\binom{n}{k} = \frac{n!}{n!(n-k)!}$$

  which is easily computed. On the other hand, listing all of the k-subsets is more difficult.

- **Search :** *Find at least one example o f a structure o f a particular type (if it exists).* A typical example of a search problem is to find a clique (A clique [Figure 1] of a graph $G(V, E)$ is a complete subgraph of $G(V, E)$) of a specified size in a given graph. Generating algorithms can sometimes be used to search for a particular structure, but for many problems, this may not be an efficient approach. Often, it is easier to find one example of a structure than it is to enumerate or generate all the structures of a specified type.

1 General Countings Methods

When one encounters a problem, one of the primary questions asked is "How to solve the problem?"

Asking questions may renge fro issues such as

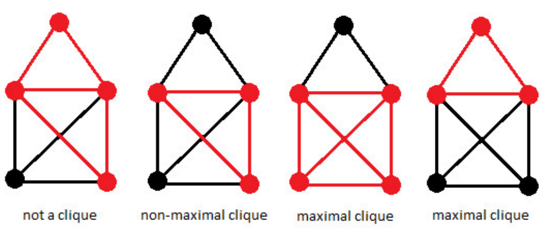


Figure 1: Clique

- What techniques to use for solving the problem ?
- How to break the problem into small manageable pieces ?
- What unseen challenges are below the surface of the problem ?
- What is the first step of the problem ?

e.g. for a counting problem, are the outcomes to be enumerated in this problem

1. unordered sets (combination)
2. ordered sets (sequences)

2 Fundamental Counting Principles

I. The Addition Problem :- If there are r_1 different objects in the first set, r_2 different objects in the second set, ...and r_m different objects in the m^{th} set, and if the **different sets are disjoint**, then the number of ways to select an object from the m-sets is

$$r_1 + r_2 + \ldots + r_m$$

.

II. The Multiplication Problem :- Suppose a procedure can be broken into m successive(ordered) stages, with r_1 diffent outcomes in the first stage, r_2 different outcomes in the first stage, ...and r_m different outcomes in the m^{th} stage.

If the number of outcomes at each stage is independent of the choices in previous stages and *if the composite outcomes are all distinct,* then the total outcomes of the procedure has

$$r_1 \times r_2 \times \ldots \times r_m$$

Note : The **Addition Principle** requires disjoint sets of objects and the **Multiplication Principle** requires that the procedure break into ordered stages and that the composite outcomes be distinct.

- The sum $(\mathbf{a} + \mathbf{b})$ is the number of items resulting when a set of $\boldsymbol{a}$ items is added to a set of $\boldsymbol{b}$ items
- The product $(\mathbf{a} \times \mathbf{b})$ is the number of sequences $\boldsymbol{A}.B$ when $\boldsymbol{A}$ can be any of $\boldsymbol{a}$ aitems and $\boldsymbol{B}$ can be any of $\boldsymbol{b}$ items

Example 2.1 Arranging Books : *There are 5 different Bengali(B) books, 6 different Hindi(H) books and 8 different Tamil(T) books. How many ways are there to pick an (unordered) pair of two books not both of same language.*

Answer 2.1

- *"Unordered" means that there is* **not** *a first book in the pair, and so as outcomes cannot be broken into first stage (first book) and a secind stage - i.e* **Multiplication Principle does not apply**.
- **Thus,** *the question is to recast or decompose this problem in such a way that the Multiplication Principle or Addition Principle can be used.*
- *The problem can be broken into* **smaller parts to which the principles apply***, which raises the question of* **What parts into which** *the problem should be broken up into*

- *If one recognises a similarity between the current problem and another previously solved problem, then the current problem is easy to solve*

The **possible cases can be** *when there can be different ways to combine the books*

Case 1) *When the 2 books consist of 1 Bengali* **and** *1 Hindi book.*

This has two stages : Pick

(a) 1 Bengali book in 5 ways **and then**

(b) 1 Hindi book in 6 ways.

$\therefore$ *the total number of ways are* $5 \times 6 = 30$ *ways*

Similarly

Case 2) *When the 2 books consist of 1 Bengali* **and** *1 Tamil :* $5 \times 8 = 40$ *ways*

Case 3) *When the 2 books consist of 1 Hindi* **and** *1 Tamil :* $6 \times 8 = 48$ *ways*

The total outcomes in all the three cases are clearly disjoint and they equate to

$$\therefore, (5 \times 6) + (5 \times 8) + (6 \times 8) = 118$$

3 Simple Arrangements and Selections

- An r-Permutation of *n-distinct* objects is an **arrangement** using *r-objects* of the *n-objects* is : $P(n,r) = \frac{n!}{(n-r)!}$
- An r-Combination of *n-distinct* objects is an **unordered selection** or **subset** *r-objects* of the *n-objects* : $C(n,r)$

Note : $C(n,r)$ are called *binomial coefficients* of the binomial expansion $(x+y)^n$

Definition 3.1 *All r-Permutations of n-objects can be generated by* **first** *picking any r-combinations of the n-objects and* **secondly** *arranging these r-objects in any order.*

$$P(n,r) = C(n,r) \times P(r,r) \tag{1}$$

$$\therefore C(n,r) = \frac{P(n,r)}{P(r,r)} = \frac{n!}{r!(n-r)!} \equiv \binom{n}{r} \tag{2}$$

Example 3.1 *How many ways are there to arrange the seven letters in the word* **SYSTEMS**.

Answer 3.1

1. *Pick the positions that the 4 letters :* **E, M, T, Y** *will occupy in the seven letter arrangements*

2. *The 3* **S***'s will fill the remaining three positions in one way.*

There are 7 possible positions for **E***, 6 possible positions for* **M***, 5 possible positions for* **T** *and 4 possible positions for* **Y**.

$\therefore$*, there are* $P(4,3) = \frac{7!}{3!} = 840$ *arrangements*

Example 3.2 *In how many of the arrangments of the above example do the 3* **S***'s appear consecutively for the word:* **SYSTEMS**.

Answer 3.2 1. *Consider the case, where the 3* **S***s appeak consecutively .*

This can be done by **grouping** *the 3* **S***'s as 1* **S** *: -* $5! = 120$ *ways*

2. *Another way to look at this problem is to think of temporarily setting aside two of the* **S***s, arranging the 5 remaining letters :* **Y, T, E, M, S** *in* $5!$ *ways.*

4 Arrangements and Selections with Repeatations

Theorem 4.1 *If there are n objects, with r_1 of* **Type 1**, *r_2 of* **Type 2**, *... and r_m of* **Type m**, *where $r_1 + r_2 + \ldots + r_m = n$, then the number of arrangements of these n objects is* [1] *:*

$$P(n : r_1, r_2, \ldots, r_m) = \binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\ldots\binom{n-r_1-r_2-\ldots-r_{m-1}}{r_m} = \frac{n!}{r_1!r_2!\ldots r_m!}$$

Proof 4.1 *Suppose that for each type, the r_i objects of type i are given subscripts numbered : $1, 2, \ldots r_i$ to make each object distinct.*

There are $n!$ arrangements of the n-distinct objects by enumerating all $P(n : r_1, r_2, \ldots, r_m)$ patterns (without subscripts) of the objects, and then placing the subscripts in all possible ways.

Thus, a pattern will have $r_1!$ ways to subscript the r_1 objects of **Type 1**, *$r_2!$ ways to subscript the r_2 objects of* **Type 2**, *... and $r_m!$ ways to subscript the r_m objects of* **Type m**. *Thus*

$$P(n : r_1, r_2, \ldots, r_m)r_1!r_2!\ldots r_m! = n!$$
$$\Rightarrow P(n : r_1, r_2, \ldots, r_m) = \frac{n!}{r_1!r_2!\ldots r_m!} \tag{3}$$

Example 4.1 *The pattern* `b a a n n a` *can have subscripts on* `a`*s as placed in the $3!$ ways*

1. $ba_1a_2nna_3$
2. $ba_2a_1nna_3$
3. $ba_3a_1nna_2$
4. $ba_3a_2nna_1$
5. $ba_1a_3nna_2$
6. $ba_2a_2nna_1$

For each of these $3!$ ways to subscript the `a`*s, similarly, there are $2!$ ways to subscript the* `n`*s*

Theorem 4.2 *The number of selections with repetations of r objects chosen from n types of objects is*

$$C(r+n-1, r) = \binom{n+r-1}{r}$$

Example 4.2 *How many ways are there to form a sequence of 10 letters from 4* `a`*, 4* `b`*, 4* `c` *and 4* `d`*, if each letter must appear at least twice*

Answer 4.1 *To apply the theorem, one must exact know* `a`*s,* `b`*s,* `c`*s and* `d`*s will be in the arrangement.*

Thus, the proble has to be broken up into subproblems that each involves sequences with given number of `a`*s,* `b`*s,* `c`*s and* `d`*s.*

There are **two categories** *of letter frequencies that sum to 10 with each letter two or more times*

Category A. *There are 4 appearances of 1 letter and 2 appearances of each other letter.*

(a) *Number of cases for* **choosing which letter** *occurs 4 times : $C(4, 1) = \frac{4!}{1!3!} = 4$* **and**

(b) *the* **number of arrangements** *for 4 of one letter and 2 of the other three letter : $P(10 : 4, 2, 2, 2) = 18,900$*

Category B. *There are 3 appearances of 2 letters and 2 appearances of each other letter.*

[1] $\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$

(a) *Number of cases for* **choosing which** *2 of the 4 letters occurs 3 times :* $C(4,2) = 4$ **and**

(b) *the* **number of arrangements** *3 of* <u>*2 letters*</u> *and 2 of the* <u>*2 others*</u> *:* $P(10 : 3, 3, 2, 2) = 25,200$

The total number of ways for arranging the 10 letter sequences are : $4 \times 18,900 + 6 \times 25,200 = 2,26,800$

Appendix (Own Interest) : Striling's Approximation Formula

For large values of n,

$$n! \approx \sqrt{2\pi n} \binom{n}{\exp}^{n}$$

Thus,

$$\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^{n}} \approx 1$$

Author's note: This article may use ideas you haven't learned yet, and might seem overly complicated. It is not. Understanding Stirling's formula is not for the faint of heart, and requires concentrating on a sustained mathematical argument over several steps.

Even if you are not interested in all the details, I hope you will still glance through the article and find something to pique your curiosity. If you are interested in the details, but don't understand something, you are urged to pester your mathematics teacher for help.

Factorials!

Unbelievably large numbers are sometimes the answers to innocent looking questions. For instance, imagine that you are playing with an ordinary deck of 52 cards. As you shuffle and re-shuffle the deck you wonder: How many ways could the deck be shuffled? That is, how may different ways can the deck be put in order? You reason that there are 52 choices for the first card, then 51 choices for the second card, then 50 for the third card, etc. This gives a total of

$$52 \times 51 \times 50 \times \cdots \times 2 \times 1$$

ways to order a deck of cards. We call this number "52 factorial" and write it as the numeral 52 with an exclamation point: 52! This number turns out to be the 68 digit monster

$$80658175170943878571660636856403766975289505440883277824000000000000$$

which means that if every one on earth shuffled cards from now until the end of the universe, at a rate of 1000 shuffles per second, we wouldn't even scratch the surface in getting all possible orders. Whew! No wonder we use exclamation marks!

For any positive integer n we calculate "n factorial" by multiplying together all integers up to and including n, that is, $n! = 1 \times 2 \times 3 \times \cdots \times n$. Here are some more examples of factorial numbers:

$1! = 1$	$2! = 2$	$3! = 6$	$4! = 24$	$5! = 120$
$6! = 720$	$7! = 5040$	$8! = 40320$	$9! = 362880$	$10! = 3628800$

Stirling's formula

Factorials start off reasonably small, but by 10! we are already in the millions, and it doesn't take long until factorials are unwieldly behemoths like 52! above. Unfortunately there is no shortcut formula for $n!$, you have to do all of the multiplication. On the other hand, there is a famous approximate formula, named after the Scottish mathematician James Stirling (1692-1770), that gives a pretty accurate idea about the size of $n!$.

Stirling's formula: $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

Before we continue, let's take a moment to contemplate the fact that n factorial involves nothing more sophisticated than ordinary multiplication of whole numbers, which Stirling's formula relates to an expression involving square roots, π (the area of a unit circle), and e (the base of the natural logarithm). Such are the surprises in store for students of mathematics.

Here is Stirling's approximation for the first ten factorial numbers:

$1! \approx 0.92$	$2! \approx 1.92$	$3! \approx 5.84$	$4! \approx 23.51$	$5! \approx 118.02$
$6! \approx 710.08$	$7! \approx 4980.39$	$8! \approx 39902.39$	$9! \approx 359536.87$	$10! \approx 3598695.62$

You can see that the larger n gets, the better the approximation proportionally. In fact the approximation $1! \approx 0.92$ is accurate to 0.08, while $10! \approx 3598695.62$ is only accurate to about 30,000. But the proportional error for $1!$ is $(1! - .92)/1! = .0800$ while for $10!$ it is $(10! - 3598695.62)/10! = .0083$, ten times smaller. This is the correct way to understand Stirling's formula, as n gets large, the proportional error $(n! - \sqrt{2\pi n}(n/e)^n)/n!$ goes to zero.

Developing approximate formulas is something of an art. You need to know when to be sloppy and when to be precise. We will make two attempts to understand Stirling's formula, the first uses easier ideas but only gives a sloppy version of the formula. We will follow that with a more sophisticated attack that uses knowledge of calculus and the natural log function. This will give us Stirling's formula up to a constant.

Attempt 1. To warm up, let's look at an approximation for the exponential function e^x. The functions $1 + y$ and e^y have the same value and the same slope when $y = 0$. This means that $1 + y \approx e^y$ when y is near zero, either positive or negative. Applying this approximation to x/n, for any x but large n, gives $1 + x/n \approx e^{x/n}$. Now if we take $n - 1$st power on both sides, we get the approximation

$$\left(1 + \frac{x}{n}\right)^{n-1} \approx e^{(n-1)x/n} \approx e^x.$$

Returning to factorials, we begin with an obvious upper bound. The number $n!$ is the product of n integers, none bigger than n, so that $n! \leq n^n$. With a bit more care, we can write $n!$ precisely as a fraction of n^n as follows:

$$n! = \left(1 - \frac{1}{2}\right)^1 \left(1 - \frac{1}{3}\right)^2 \cdots \left(1 - \frac{1}{n}\right)^{n-1} n^n.$$

I won't deprive you of the pleasure of working out the algebra to confirm that this formula is really correct. Using the approximation for the exponential function e^x we can replace each of the factors $(1 - 1/k)^{k-1}$ by e^{-1} and arrive at $n! \approx e\,(n/e)^n$. Because of cumulative errors, the formula $e\,(n/e)^n$ sorely underestimates $n!$, but it does have the right order of magnitude and explains where the factor "e" comes from.

Attempt 2. Mathematically, addition is easier to handle than multiplication so our next attempt to get Stirling's formula converts it into an addition problem by taking logs. Our warmup problem this time is an approximate formula for the natural log function. We start with the series expansion

$$\frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots.$$

Substitute $x = 1/(2j+1)$ and rearrange to get

$$\left(j+\frac{1}{2}\right)\ln\left(1+\frac{1}{j}\right) - 1 = \frac{1}{3(2j+1)^2} + \frac{1}{5(2j+1)^4} + \frac{1}{7(2j+1)^6}\cdots$$

Now replacing the sequence of odd numbers $3, 5, 7, \ldots$ by the value 3 in the denominator makes the result bigger, so we have the inequality

$$\left(j+\frac{1}{2}\right)\ln\left(1+\frac{1}{j}\right) - 1 \leq \frac{1}{3}\left(\frac{1}{(2j+1)^2} + \frac{1}{(2j+1)^4} + \frac{1}{(2j+1)^6} + \cdots\right)$$

The sum on the right takes the form of the famous "geometric series"

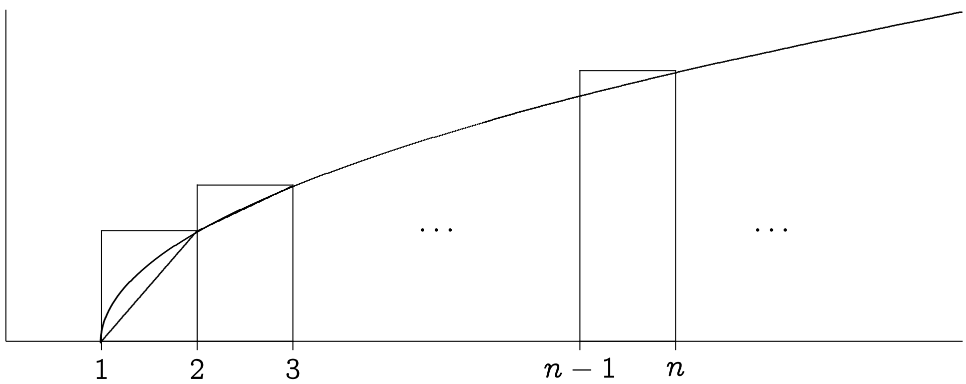
$$\rho + \rho^2 + \rho^3 + \cdots = \frac{\rho}{1-\rho}.$$

Making the replacement $\rho = 1/(2j+1)^2$ and a little algebra yields

$$\left(j+\frac{1}{2}\right)\ln\left(1+\frac{1}{j}\right) - 1 \leq \frac{1}{3}\left(\frac{(2j+1)^{-2}}{1-(2j+1)^{-2}}\right) = \frac{1}{12}\left(\frac{1}{j} - \frac{1}{j+1}\right). \tag{1}$$

All that work to show that $(j + 1/2)\ln(1 + 1/j) - 1$ is pretty close to zero. If you are inclined, you could program your computer to calculate both sides of (1) for various values of j, just to check that the right hand side really is bigger than the left. Note that we have an upper bound in (1), instead of an approximate formula. This means that the values on the two sides are not necessarily close together, only that the value on the right is bigger.

You will be relieved to hear that we are finally ready to return to Stirling's approximation for $n!$. Taking the natural log on both sides of $n! = 1 \times 2 \times \cdots \times n$, turns the multiplication into addition: $\ln(n!) = \ln(1) + \ln(2) + \cdots + \ln(n)$. This sum, in turn, is the area of the first $n-1$ rectangles pictured here. The curve in the picture is $\ln(x)$, and it reminds us that $\ln(1) = 0$.



The area of each rectangle is the area under the curve, plus the area of the triangle at the top, minus the overlap. In other words, using the definitions below we have $r_j = c_j + t_j - \varepsilon_j$.

$$\begin{aligned} \text{rectangle} &:= \quad r_j = \ln(j+1) \\ \text{curve} &:= \quad c_j = \int_j^{j+1} \ln(x)\,dx \\ \text{triangle} &:= \quad t_j = \frac{1}{2}[\ln(j+1) - \ln(j)] \\ \text{overlap} &:= \quad \varepsilon_j = \left(j + \frac{1}{2}\right) \ln\left(1 + \frac{1}{j}\right) - 1 \end{aligned}$$

The overlap ε_j is a small sliver shaped region that is barely visible in the picture, except in the first rectangle. Using the inequality (1) we worked so hard to establish, we add up on both sides and see that the infinite series satisfies $\sum_{j=n}^{\infty} \varepsilon_j < 1/(12n)$, for any $n = 1, 2, 3, \ldots$.

To approximate $\ln(n!) = \sum_{j=1}^{n-1} r_j$, we begin by splitting r_j into parts

$$\ln(n!) = \sum_{j=1}^{n-1} c_j + \sum_{j=1}^{n-1} t_j - \sum_{j=1}^{n-1} \varepsilon_j.$$

Since $\sum_{j=1}^{n-1} c_j$ is an integral over the range 1 to n, and $\sum_{j=1}^{n-1} t_j$ is a telescoping sum, this simplifies to

$$\begin{aligned} \ln(n!) &= \int_1^n \ln(x)\,dx + \frac{1}{2}\ln(n) - \sum_{j=1}^{n-1} \varepsilon_j \\ &= n\ln(n) - n + 1 + \frac{1}{2}\ln(n) - (\sum_{j=1}^{\infty} \varepsilon_j - \sum_{j=n}^{\infty} \varepsilon_j). \end{aligned}$$

Taking the exponential gives

$$n! = e^{1 - \sum_{j=1}^{\infty} \varepsilon_j} \sqrt{n} \left(\frac{n}{e}\right)^n e^{\sum_{j=n}^{\infty} \varepsilon_j}$$

Pause to note that this is an exact equation, not approximate. It gives $n!$ as the product of an unknown constant, the term $\sqrt{n}\,(n/e)^n$, and a term $e^{\sum_{j=n}^{\infty} \varepsilon_j}$ that converges to 1 as $n \to \infty$. With the formula $\sum_{j=n}^{\infty} \varepsilon_j < 1/(12n)$ we now have the bounds

$$C\sqrt{n}\left(\frac{n}{e}\right)^n \leq n! \leq C\sqrt{n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}},$$

where $e^{\frac{11}{12}} \leq C \leq e$. Once we've identified $C = \sqrt{2\pi}$, we get

$$\boxed{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}}$$

If you've made it this far, congratulations! Now you see why Stirling's formula works. The part we skipped, to show that the unknown constant C is actually equal to $\sqrt{2\pi}$ is not an easy step. But we leave this aside, and look at some other properties of the number $n!$.

Number of digits For any $x > 0$ the formula $d(x) = \lfloor \log_{10}(x) \rfloor + 1$ gives the number of digits of x to the left of the decimal point. For moderate sized factorials we can simply plug this formula into a computer to see how many digits $n!$ has. For example, $d(52!) = 68$ and $d(1000000!) = 5565709$. But suppose we wanted to find the number of digits in a really large factorial, say googol factorial? (Googol means ten raised to the power 100 or 10^{100}). Even a computer can't calculate googol factorial, so we must use Stirling's formula. Let $g = 10^{100}$, substitute into Stirling's formula, and take log (base 10) on both sides to obtain

$$\log_{10}(\sqrt{2\pi g}\left(\frac{g}{e}\right)^g) \quad \leq \log_{10}(g!) \quad \leq \log_{10}(\sqrt{2\pi g}\left(\frac{g}{e}\right)^g e^{1/12g}).$$

Let's concentrate on the left side $\log_{10}(\sqrt{2\pi g}(g/e)^g)$. Using the logarithm property and the fact that $\log_{10}(g) = 100$, we simplify this to $\log_{10}(\sqrt{2\pi}) + 50 + g(100 - \log_{10}(e))$. The hard part of this calculation is to find $\log_{10}(e)$ to over 100 decimal places, but the computer is happy to do it for us. Once this is accomplished we find that

$$\begin{aligned}\log_{10}(\sqrt{2\pi g}(g/e)^g) \; &= 995657055180967481723488710810833949177056029941963334 33\\ &\,8855462168341353507911292252707750506615682567.21202883\ldots\end{aligned}$$

When we knock off the decimal part and add 1, we get $d(\sqrt{2\pi g}(g/e)^g)$. We can be sure that the number of digits in googol factorial is the same by comparing with the upper bound. The right hand side $\log_{10}(\sqrt{2\pi g}(g/e)^g e^{1/12g})$ exceeds the left hand side only by the minuscule amount $\log_{10}(e^{1/12g}) = \log_{10}(e)/12g$. When this is added to the fractional part .21202883..., the first hundred or so digits after the decimal point are not changed. Therefore $d(\sqrt{2\pi g}(g/e)^g e^{1/12g})$ is equal to $d(\sqrt{2\pi g}(g/e)^g)$, and since $d(g!)$ is in between, it also must be the same.

Raising 10 to the power of the fractional part .21202883... gives us the first few digits of $g!$, so we conclude that googol factorial is $g! = 16294 \cdots 00000$, where the dots stand in for the rest of the exactly

$d(g!)$=99565705518096748172348871081083394917705602994196333433885546216834135350791129225270775050661568 2568

digits. This explains why no one can or ever will calculate all the digits of googol factorial. Where would you put it? A library filled with books containing nothing but digits? A trillion trillion computer hard drives? None of these puny containers could hold it. This super-monster has more digits than the number of atoms in the universe.

Trailing zeros Looking back, you may notice that 52! ends with a stream of zeros. For that matter, all the factorials starting with 5!, have zeros at the end. Let's try to figure out how many zeros there will be at the end of $n!$. This doesn't rely on approximate values of $n!$ anymore, more importantly we need to understand the divisors of $n!$.

Each zero at the end of $n!$ comes from a factor of 10. For instance, 10! has two zeros at the end, one of which comes from multiplying the 2 and the 5.

$$\begin{aligned} 10! &= 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \\ &= (1 \times 3 \times 4 \times 6 \times 7 \times 8 \times 9) \times (2 \times 5) \times 10 \\ &= (36288) \times (100) \end{aligned}$$

The fact that 36288 is an even number means that there are extra factors of 2 that don't get matched with any 5's. Since there is always an excess of 2's, the number of trailing zeros in $n!$ is equal to the number of 5's that go into $n!$.

Imagine lining up all the numbers from 1 to n to be multiplied. You will notice that every fifth number contributes a factor of 5, so the total number of 5's that factor $n!$ should be about $n/5$. Since this isn't an integer, we knock off the fractional part and retain $\lfloor n/5 \rfloor$.

$$\begin{array}{cccccccccccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & \cdots & n \\ & & & & \uparrow & & & & & \uparrow & & & & & \uparrow & & & & & \uparrow & & & \end{array}$$

According to this formula, the number of trailing zeros in 10! is $\lfloor 10/5 \rfloor = 2$, and that checks out. But for 52! the formula gives $\lfloor 52/5 \rfloor = 10$, when there are really 12 trailing zeros. What's going on? The problem is that we forgot to take into account that the number 25 contributes *two* factors of 5, and does 50. That's where the extra two zeros come from.

Now we modify our formula for the number of trailing zeros in $n!$ to

$$z(n) = \lfloor n/5 \rfloor + \lfloor n/25 \rfloor + \lfloor n/125 \rfloor + \lfloor n/625 \rfloor + \cdots$$

We have anticipated that all multiples of 125 give three factors of 5, multiples of 625 give four factors of 5, etc. Also note that if n is less than 25, for instance, then the formula $\lfloor n/25 \rfloor$ automatically returns a zero.

We can get an upper bound on the number of zeros by not knocking off the fractional part of $n/5^j$ and using the geometric series

$$z(n) = \sum_{j=1}^{\infty} \left\lfloor \frac{n}{5^j} \right\rfloor \leq \sum_{j=1}^{\infty} \frac{n}{5^j} = \frac{n}{4}.$$

This turns out to be pretty close to the right answer. In other words, the number of trailing zeros in $n!$ is approximately $n/4$. For example, the number of trailing zeros in googol factorial works out to be exactly $z(g) = g/4 - 18$ or

$$\begin{array}{c} 24999999999999999999999999999999999999999999999999 \\ 99999999999999999999999999999999999999999999999982. \end{array}$$