

FULL PROBABILITY MODEL FOR HUMAN GENETIC VARIATION

ZULFIKAR MOINUDDIN AHMED

1. HUMAN GENETIC VARIATION

In the past few weeks, I had been concentrating on understanding Human Genetic Variation in a concrete manner. This would never have been a project that interested me five years ago. Five years ago, I had deep dissatisfaction with nature of arbitrage-free movements of the price levels of assets in liquid capital markets, on the classical field theory, Four-Sphere Theory, and the problems with work of Einstein and Schroedinger. The eventual success of those projects led me to feel satisfied that I have some answers in Finance and Fundamental Macroscopic Physics, and I had turned to Personality Psychology and other issues of Social Science. For many years I had been developing my own convictions regarding a Single Human Race. For some years I had known that Sewall Wright's index of fixity did not allow a great difference across continental groups genetically compared to within groups.

Finally in the last couple of weeks at most, especially after my great success vindicating Aristotle's basic thesis that higher moral virtues lead to eudaimonia proxied by modern Life Satisfaction metrics led me to consider concretely the the probability model of Human Genetic Variation directly.

2. THE 1000 GENOME PROJECT PUBLIC DATA

Let us return to the elementary description of what is known about Genetic Diversity in humans. In around 6 billion base pairs that are fixed, there are 0.1% locations of the human genome where multiple DNA letters are possible. Thus a full probability model for human genetic variation is just a long list of probabilities of various possibilities at only those locations. These locations are called snip locations. I am now attempting to extract a long list of probabilities of alleles (differing DNA letters) at all the snip locations of the human genome.

I am not going to be doing all sorts of sequencing myself, so I take advantage of public services made available by the 1000 Genome Project.

Using their facility to determine snip locations [1], I create a large VCF file. Then I download it and use the python library PyVCF to extract allele frequencies.

```
import vcf
vcf_reader = vcf.Reader(filename='all.vcf.gz',encoding='utf-8')
for rec in vcf_reader:
    print(str(rec.POS)+", "+str(rec.REF) +
          ", " + str(rec.ALT) + ", " + str(rec.INFO['AF'][0]))
```

Then I wait perhaps for some days till I have all the allele frequencies for Chromosome 1.

Date: June 15, 2021.

3. GOOGLE DRIVE FOR PUBLIC AVAILABILITY

I intend to produce allele frequencies for the entire Human Genome in the next weeks and make them available publicly because these are universal scientific data. These data, in toto, is the complete description of human genetic variation, not just for the 7.8 billion humans alive but also all possible human beings that could occur at all. There are some exceptional situations but this is almost complete truth.

So with all the concrete allele frequencies and snip locations we have all the data to understand concretely human genetic diversity.

4. HUMAN RACE IS A SINGLE RACE

Carolus Linnaeus, Swedish taxonomist extraordinaire, produced a taxonomy of living things long before Darwin. He lived between 1707-1778. On human beings, he classified them into five 'races'. People often do not have clear perspective by my experience, so I remind the reader that before Francis Crick and James Watson's discovery of double helix structure of genetic code of living things in 1953, which is 175 years after the death of Linnaeus, there was very little strongly objective to either confirm or deny Linnaeus' taxonomy of human 'subspecies'. He was a botanist primarily and so for him 'subspecies' was a natural sort of category. It is a disaster for the human race that his speculative division simply became stuck in the popular culture and even today there is a vast confusion about human diversity. My position for the past decade has been that we cannot be indifferent to whether there is any fundamental sense in these sorts of classifications. Being wrong has vast and tragic consequences. Economist recently did a clever article on the impact of a movie by the Ku Klux Klan, a propaganda film called "The Klansman: The Birth of a Nation". It was shown in theaters in America in the 1920s. The Economist found that in the cities and towns where it was shown, there was an upsurge of lynchings of black people, and these did not stop till 1942. In other words classification schemes for human beings are not innocent and objective but have tragic consequences when there is some success in dehumanising some people and the mobs are quite bloodthirsty when their savagery is unleashed with some conviction that they are on the right. These shenanigans must end. We must learn to understand clearly the objective facts regarding our beloved people the Human Race.

Indeed it is rather difficult to justify multiple human 'subspecies'. I have come across many undereducated people who believe that multiple subspecies exist among human beings. They are primitive, savage, and ignorant people, in a way, but it is mostly the irresponsibility of the authorities and leaders that are responsible for this dreadful state of the world.

What is the objective situation? Well, there exists a positive integer, N_{snips} and a set of probabilities $0 \leq p_r \leq 1$ for $r = 1, \dots, N_{snips}$ such that, roughly the $2^{N_{snips}}$ binary states give us the full spectrum of human genetic diversity possible for human beings. There are many ways of producing subsets of all human beings. Ethnic tribalists are seeking acceptance of their classification scheme. They better have much much better justification for their schemes than they have produced to date, for their justification are not compelling to any rational and sensible human being.

I offer an alternative view, which is that there is a far greater justification to divide up the human race for the *benefit of life satisfaction of all people and higher*

civilisation of human race, and to destructively denigrate tribal divisions so that the future is better for all human beings.

5. BILL GATES IS UNHAPPY WITH INHUMANE TREATMENT HE GETS

Bill Gates has conviction that African-American slaves were treated in a humane manner and he is unhappy with the inhumane treatment he receives. I would suggest a model African-American owned cotton plantation where Bill Gates could be the slave. He would be treated humanely by his masters and can pick cotton everyday in a utopian situation and be treated so well that he would never want his freedom at all.

In a pastoral, idyllic and tranquil Southern cotton plantation owned by African-American masters, Bill Gates will be a happy slave without a worry in the world, and Southern economy will improve drastically because Bill Gates will be such a good slave that he can single-handedly produce a superconducting slave economy by his own great work as a cotton-picking slave! We can confiscate all his property and let him prove his convictions in this way.

6. OBVIOUSLY I AM GENETICALLY SUPERIOR TO BILL GATES

It's true that the Human Race is a Single Race, and I do not believe that Bill Gates is a *different race* than myself. Both he and I are in the same Human Race. My genetic superiority to Bill Gates is not a *racial* matter. It has to do with our actual ancestries. You see, Bill Gates' grandfather sold furniture, and his ancestors around 0 AD were not even literate in England. In contrast both my late father K. M. Ahmed and his before him, Mr. Jamal-uddin Ahmed, were highly literate people who had served the people of Bengal in government and my illustrious ancestry reaches back to Ancient Vedic nobility in Northwest India, urban literary culture already in 1750 BC and Vangala of 500 BC was a sophisticated Civilisation when Bill Gates' ancestors were worshipping various sorts of stones and were foragers.

7. STATISTICAL DEPENDENCE OF SNIPS

I am waiting right now for all the allele frequencies for Chromosome 1 to be computed. It is quite clear that vectors of length N_{snips} are too much for human genetic diversity. In other words the truth about human diversity will require us to take into account the statistical *dependence* of variations at the snip locations. We have therefore a new technical problem to resolve, i.e. the determination efficiently of the dependence structure of several million variables without combinatorial algorithms that are intractable.

My idea here is to divide up snip locations with very low probability minor alleles, and make the assumption that these cases are *determined completely* by other locations.

Now let us say $N_{snips} = 6 \times 10^6$. Then $2^{N_{snips}}$ is too large to be a realistic value for the actual constraint on the total possible human beings. This means of course that the independence hypothesis is not realistic, and the actual fraction of possible human beings is vastly smaller than roughly $10^{N_{snips}/3}$.

8. WHAT DETERMINES AUTHORITY ON TRUTH?

Is it right to hold views against the grain of the entire Age? If you ask Friedrich Nietzsche, the answer would be that it is not only right but a duty to yourself to hold independent views. Ralph Waldo Emerson would say something similar from his 1841 essay, "Self-Reliance". This is one side of the question. On the flip side of the question is the sort of 'independence' that Bill Gates shows, the self-justification for malevolence towards people.

I think the answer ought to be that there is truth, and justice, and great and generous heart, and virtue, and there is knowledge and ignorance. We have faculties to judge ourselves, we have ability to learn and correct our own errors. The substantial issue is not whether we ought to be rebels or whether we should conform to others' demands. The substantial issues are whether we are right about things and enlightened and just regardless of whether others consider us rebels or venerate us as authorities. Those things are social events and truth is deeper than all that, and truth is also quite difficult to discern. I am Naturalist in my view about Human Race. I hold strongly now the view that Human Race is an Angelic Race that is young and still has much to learn about how to produce a good civilisation. We have produced great progress in the last twelve thousand years since civilisations began forming in the Middle East first, with Mesopotamia, then Egypt and China and India and then progress to this very day. We will live once and not know when we will return, or even if we will return at all, and so we ought to have responsibility to ourselves to ensure that we are comfortable with our conscience and our senses, and so external noise does not intrude upon what we must do with our lives.

I have been looking at World Values Survey moral values across the world and it is quite clear to me that Conscience is human nature, even if Reason is not. Of course there are massive failures like the Psychopath Bill Gates but they are a tiny set of exceptions in the human race along with serial killers like Charles Manson.

9. DISCRETE CORRELATION CAN PRODUCE TRACTABLE DEPENDENCE INFERENCES

With actual samples rather than just the allele frequencies at different locations, we can use my discrete correlation metric to determine the dependence structure at snip locations. And this is a matrix computation overall. It looks like VCF files can contain sample information, although I do not know if the VCF files I was using have them.

The algorithm is clear from samples. We create 0-1 vectors of length N_{snip} and calculate the discrete correlations, and then we can just reduce N_{snip} based on dependence structures because we are interested in uniqueness; it could be that there are entire subsets of N_{snips} that move together and this could reduce the dimension for human genetic variation.

The main advantage of discrete correlation matrices is the simplicity of the algorithm and the potential for determining the complete dependence structure by a single matrix operation of all columns against all columns. Then we can decompose the sets of indices into mutually independent blocks and obtain $N_{blocksnips}$. I am rather optimistic that we can have a reduction of state space to size $2^{N_{blocksnips}}$ with $N_{blocksnips} \ll N_{snips}$.

10. NIETZSCHE'S GREAT ERROR

Nietzsche was quite bright, but his elitism for whatever reason simply assumed that people are literally following and molding themselves to the dictates and leadership of other people. Thus he wrote *Genealogy of Morals*. Now this is a truly marvelous book but I have looked in the past year at moral values of billions of people, whose regularity cannot sustain Nietzsche's presumption that some Aristocrats artificially concocted moral values that others followed. This is what he saw because he was not as steeped into evolutionary biology. Darwin was relatively new when he was alive, and there was no DNA at all. Today, it is easy to understand that morals are regular across the globe because there are adaptive pressures. It was only in 2008 that a Japanese team found neural correlates of moral judgment. Nietzsche's notion that one could create arbitrary values is invalid. He had wonderful insights but his notions were based on the idea that he would overturn the Christian order all by his own efforts. He was simply *wrong* about the way in which human beings are moral. It's not a permanent error; his genius is still extraordinary nonetheless even if his *Genealogy of Morals* was not the full and true story of human moral nature.

11. BILL GATES' EFFORTS TO PRODUCE A WHITE RACIAL ORDER IS DOOMED

It is not just my beautiful work that shows that Ethnic effects on Human Moral Values is bounded between 6.5-9.5%. He has only \$130.5 billion to move the world opinions when there is a 78% global consensus on importance of securing Human Rights. Bill Gates is truly a *tragic figure* in the classical sense. His hubris is spectacular, but his intelligence is quite limited and his education and cultivation is non-existent. These are the elements of tragedy. Now I would be far more sympathetic to the tragic fate of the man if I were not directly attacked by his malevolence. Zulf is quite benevolent generally, but he has irritated me, and I believe he deserves to be relentlessly bombed and physically obliterated. So having lost my sympathy, he is doomed.

12. OPEN ACCESS OF ALLELE FREQUENCIES OF CHROMOSOME 1

These [2] are the allele frequencies for locations up to index 250 million of human chromosome 1. In the next weeks my open access site will have the full allele frequencies of the entire human genome.

REFERENCES

- [1] http://grch37.ensembl.org/Homo_sapiens/Tools/AlleleFrequency?db=core
- [2] <https://drive.google.com/drive/folders/1TDYIKDK35Pve4adIJcVIN-23-QPqcacg?usp=sharing>