Lecture 0: Appetizer

Probability theory is just as simple as counting. But counting is very hard! When we rigorously apply the tools of probability to data, we get indisputable statistics, but what exactly the statistics don't dispute often remains a mystery only known to God. Like the rest of math, probability reminds me of Alexandr Block's Demon:

Behind me you must go, behind me,
My slave obedient and true;
The sparkling mountain-ridges find me
In flight unfaltering with you.

Above abysses I shall take you,
Bottomless pits of mystery;
And there, while futile terrors shake you,
Is inspiration's strength for me.

Amid the ether's flaming shower
I do not let you swoon, but show
My shadowy wings and sinewy power
To lift you and not let you go.

Upon the hills in white resplendence,
Upon the unstained meadow-ground,
In beautiful divine attendance
My fire shall strangely burn around.

Know you how frail is that delusion
By which mankind is tricked, how small
Is the poor pitiful confusion
That by wild passion's name we call?

When shadows gather in the even
And my enchantment senses you,
You wish to fly aloft to heaven
Through fiery deserts of the blue.

I gather you in my embraces
And raise you up with me afar
To where a star is like earth's places
And earth's not different from a star.

Then stricken dumb with admiration,
New universes you can see,
Sights unbelievable, creation
Made by my playful fantasy.

In fear and strengthlessness you shiver;
I hear you whisper: «Let me go!»
You from my soft wings I deliver
And smile upon you, «Fly below!»

Beneath my smile divinely winning,
In an annihilating flight,
Like a cold stone, you flutter, spinning
Into the glittering void of night.

https://www.youtube.com/watch?v=Yt6EinZBoks

How can counting - which even toddlers can perform on their fingers - create such strange experience? Let us look at a few simple examples:

1. You are invited to a household known to have 2 children. You have never seen these children and have no prior knowledge about them. As you park your car, you notice that one child is playing in the yard and that she is a girl.

   a) What is the probability both children are girls?

   Answer: Since you don't know if you are observing the youngest or oldest child, there are 3 equally likely possibilities (g, g), (b, g), (g, b), where the first coordinate stands for oldest child and the next coordinate represents the youngest child. Thus the desired probability is 1/3.

b) Her mother comes to greet you and as she yaps away you suddenly hear "...Oh this child was born on Wednesday...". What is the probability that both children are girls?

Answer: It is a mistake to think that the gender of the children is the only relevant information. If we assume that each child is equally and independently likely to be born a boy or a girl and is equally and independently likely to come into existence on any given day of the week, then the relevant objects are vectors of the form (gender of oldest, the day of its birth, gender of youngest, the day of its birth) = (x, t, y, s). The given information tells us that we either have the case (g, w, y, s) or (x, t, g, w). Each of the two vectors represents a total of $2 \times 7 = 14$ possibilities with only one overlap, namely (g, w, g, w). Thus we are faced with a total of $2 \times 14 - 1 = 27$ possibilities, of which vectors of the form (g, w, g, s) and (g, t, g, w) represent outcomes in which both children are female. The event (g, w, g, w) is the only one in common. Thus, there are $2 \times 7 - 1 = 13$ possibilities. Hence the desired probability is 13/27

Now let us consider how incorrect perceptions of probability can affect our estimations of risk and efficacy of medical intervention.

2. Consider a hypothetical scenario in which 1 in 1 000 people has a certain disease, and estimate the probability of disease after a positive and negative result of a test with sensitivity of 100% and a specificity of 95%. Sensitivity is the probability that a person with the disease will have a positive test result. Specificity is the probability that a person without the disease will have a negative result.

Answer: We shall study this phenomena in detail as we delve into conditional probability later in the semester. For now I will merely declare the results without explaining them. (If you cannot wait, we can discuss the results after class). If patient tests positive, his probability of having the disease comes to 1/50.95 or about 2%. If the patient tests negative, then he doesn't have the illness.

3. By some accounts, 50% of COVID related hospitalizations in Israel are currently among fully vaccinated individuals. If 85% of the adult Israeli population is vaccinated, does this information support efficacy of the vaccine?

Answer: If this information is as stated, a calculation shows that vaccine efficiency for preventing hospitalization is at 82%. We can examine such calculations in detail once we get familiar with conditional probability.

4. Does the 82% estimate of vaccine efficiency suggest that getting vaccinated is the right choice?

Answer: One dimensional analysis can only suggest optimizations in one direction, which are not globally optimizations at all. Taking blood thinners, for instance, will likely reduce the chance of thrombosis. Does this mean that infants and 20 year olds should be on blood thinner medications? At the extreme end, a bag over your head will surely make the chance of a COVID related medication nill.

If you are interested in COVID related risk analysis, I highly suggest the YouTube channel "Ivor Cummings" (specifically the discussion with the Nobel prize laurite Prof. Michael Levitt) and the book "COVID: Why most of what you know is wrong" by Sebastian Rushworth. Dr. Rushworth is a Swedish physician who has treated COVID patients and has a much closer acquaintance with statistical inference than the common medical practitioner.