Today’s student faces an onslaught. We’ve created technology under the guise of absolute knowledge and readily deny the many errors and bugs that refute the reality of this absolute desire. At the center of this information geyser surely is the hypothesis test, the way in which we are able to check our intuition with precise logical and mathematical detail. However, modern mathematics rarely claims absolute truth. Methods, such as the confidence interval, the posterior distribution, and the standard normal distribution, handle our degree of belief, not belief itself.

The problem students face is that truth is projected into a server, not into the Self. The solution of which is not to abolish the server, or hide it underground, or obfuscate its physical reality into the clouds, but to recenter the computational burden back to the Self. Many scientists recenter this belief by asking questions that do not yet have answers. As a Yogi, meditative absorption on the Self is infinite, multi-faced, and in union with the physical world, not in conflict with it.

Predicting the occurrence of a heart attack is difficult. Our current diagnostic model comes largely from the Framingham Heart Study, which began as an investigation into the epidemiology of cardiovascular disease in about 5000 people. The study continues to this day with more people and continued investigations on the molecular epidemiological basis of cardiovascular disease. Clinically, we are interested in being able to predict the risk of having a myocardial infarction in a patient. The following are the primary measures of predicting risk in a patient, based on this study:

Age
Sex
Smoking Status
Total Cholesterol
HDL Cholesterol
Systolic Blood Pressure
[Other potential markers]

With this set of relative risks, we are trying to calculate the probability of a cardiovascular event in 10 years, 20 years, 30 years etc. We know a lot about the physiology of the heart and what happens in a heart attack, but predicting one seems to be difficult in any absolute, definitive way. We still seem to be playing dice, something I need to elaborate on.

Of the half-million bikes in London on any given day, let's say that approximately 1000 of those bikes’ chains fall off at some time that day. So you get up in the morning and hop on your bike. What is the chance that your bike’s chain will fall off? If you thought 1000/0.5 million or 0.2%, you are wrong. Having a bike chain fall off is dependent on many factors, all of which determine if the chain will fall off. For instance, if you inspected your bike before you got on and noticed that a couple of teeth in your chainring had broke off, the probability that the chain will fall off that day is definitely higher than 0.2%.

So let's now reform our analogy. Let's say that we now remove all bikes who's chainrings are missing teeth from our initial survey. This probably wouldn't change the absolute number of chains that fell off by much, unless there has been a huge recent problem in chainring construction. So, our proportion of fallen chains to total number of bicyclists is maybe 0.19%. Now you go to get on your bicycle and inspect the chainring and everything looks in order. What is the probability that your chain will fall off now? The right answer is "less than if I was missing teeth in my chainring, but my individual chance of having my chain fall off is dependent on several factors that have to do with the theoretically completely predictable mechanism of my bicycle." Now that I am so focused on my chain not falling off, I may also behave differently on my bike. I may add a little oil, clean up some dirt I notice in my derailer, or even dodge that pothole to keep myself safe.

If you haven't made the connection yet, your heart is the bike. There are predictive mechanisms that lend a doctor to suspect that it will infarct and of a certain number of people in your exact circumstance, a proportion of them have an adverse cardiac event. Does that mean your individual chance is that proportion? The problem with this question is that we cannot say that each individual in a population is physiologically the same enough to lend the outcome to pure chance. As clinicians, this is all that can be done in accord with best practices. The public health and epidemiological evidence allows us to make population based decisions about how to treat a group of individuals, not the individual itself. The additional information we gain about our own heart then becomes psychological ammo to help us come to terms with our own confidence in whether we feel we will or will not have a heart attack. It becomes an intuitive understanding.

The advantage of having a clinical conversation about our hearts adds a substantial amount of detail to our understanding of ourselves and how we will make decisions. For instance, to eat pie in light of high cholesterol requires a certain level of denial, but what if it was made by a great mother? Wouldn't that be therapeutic for someone facing their own death with every bit of food?

How can we step closer with a clear and non judgmental mind and explore and evaluate our own individual situations in not quite a fully objective way, but one that is the most honest and therapeutic to those facing death. Is the affection toward those that suffer meaningless? Is there a key, pattern, or set of mechanistic relationships that can allow my logical frontal cortex to comprehend the biochemical basis of heart disease? Or is it the limbic system and midbrain, and those old, conserved bundles in the arcuate nucleus that provide the intuitive capacity to act, to further deny the call to death.

It seems then that our task is not to ask if there is a psychological component to not only our study of cardiovascular and metabolic disease and its complication, but also to the inherent bias our classification schema has produced in ourselves. A shift in our understanding of probability and objectivity may in fact reveal something about ourselves and our desires.

The problem arises in our scientific consciousness however, as Jung said, “It’s only psychological often means ‘it’s nothing’.

What we eat affects our longevity and health. Mechanistic molecular investigations have honed our understanding of many important micronutrients. Large scale epidemiology studies about nutrition also seem to indicate a large influence of our dietary choices on our health. However, the guilt and anxiety about what we put in our bodies seems to be put aside. I wish we had a better way to represent the subjective experience of available data that allowed us to honor and accept who we are now, yet let us move on to happier and healthier lives.

“We know how to protect ourselves from physical contagion, but do not yet bother about psychic contagion, even though we witness regularly its destructive effects. Jung cautions against mass solutions,” (as does Fisher!) “which only reinforce a lack of individual responsibility, and invites us to do the simplest thing: take ourselves more seriously. Interest in [human] nature would not lead to increased self-absorption, but would encourage the natural maturation process whereby people can form a stable communitus.” - Meredith Sabini Ph.D. “The Earth Has a Soul”

It seems we need to take a journey to the depths of probability, mechanism, and the nature of reality itself. Because this is about the tension between objectivity and subjectivity, determinism and probability, perhaps maybe this echos old debates between theorists and experimentalists, Plato and Aristotle.

<u>Single and Multiple Hypothesis Testing</u>
From Bradley Efron’s “Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction”

False Discovery Rate $\phi(z)$ or $Fdr(z)$ (1)

for gene i H_0: gene i is "null"

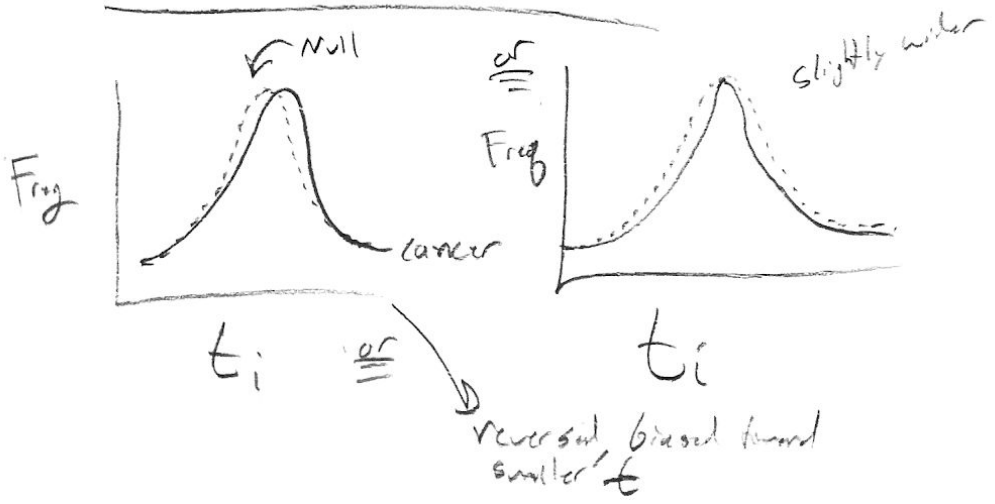
Let x denote the gene's expression level (TAG signal)

For a single patient:

t-test does not reject H_0!

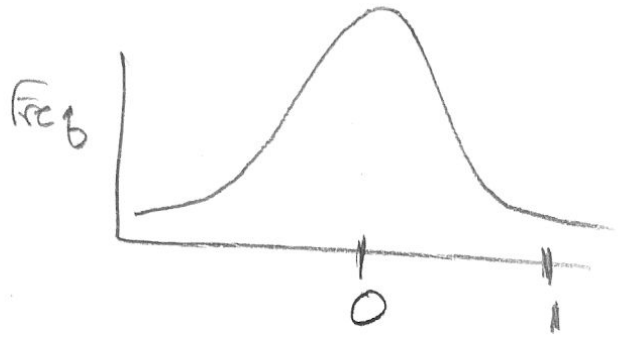
$$t_i = \frac{\bar{x}_{i\,\text{cancer}} - \bar{x}_{i\,\text{Normal}}}{s_i}$$

If t-test rejects H_0



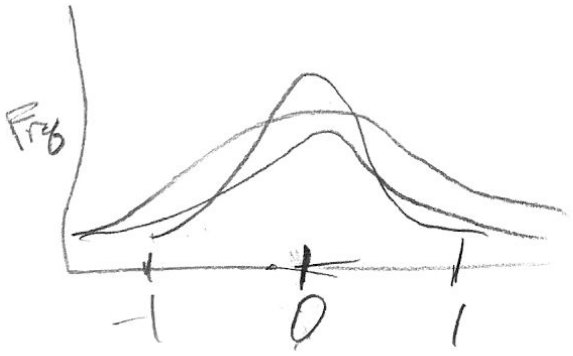
Lets continue to assume the expression levels for cancer and non-cancer are more-or-less Normal

Take the mean, $\bar{x} = 0$
and the standard deviation, $s = 1$



where $Z = \frac{x_i - \bar{x}}{s}$

Bayesian statistics — The true state of the world as expressed by mathematical statistics, is expressed by the degree of belief or Bayesian probability.

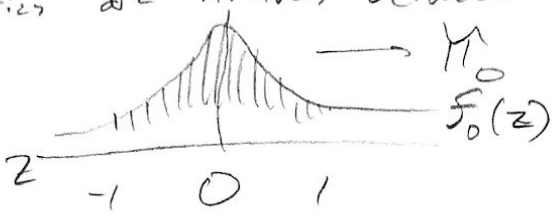


lets call

$\pi_0 = Pr\{\text{null}\}$ of $f_0(z)$ = density if null

$\pi_1 = Pr\{\text{non-null}\}$ of $f_1(z)$ = density if non-null

Probabilities are numbers between 0 – 1!



* could draw another for $f_1(z)$ and π_1 for non-null

* The purpose of Large-scale hypothesis testing is to reduce vast collections of possible important variables down to only the most meaningful.

For this to be true our null distribution must contain the majority of non-significant variables, such as $\pi_0 \geq 0.90$

Using the standard normal distribution we can use a mathematical formula to express the distribution and thus calculate our probabilities by integration (with boundary conditions).

From table rapidtables.com $f_0(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ for z dist: $\sigma = 1$ and $z = \frac{x_i - \bar{x}}{s}$ and $s = \sigma$

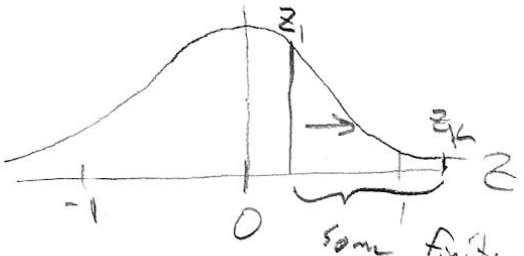
$$f_0(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(z)^2}{2}} = \boxed{\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}}$$

Mixture density and mixture probability distribution

For a continuous random variable, z
the probability density function is $f(z)$

* The probability <u>distribution</u> function is over and subset $z_1, \ldots z_K$, called Z

$$\int_{z_1}^{z_K} f_0(z)\,dz = F_0(Z)$$



some finite probability, it's calculatable!

same for $f_1(z)$, our non-null distribution

$$F_1(Z) = \int_Z f_1(z)\,dz \quad \text{with } Z = z_1, \ldots z_K$$

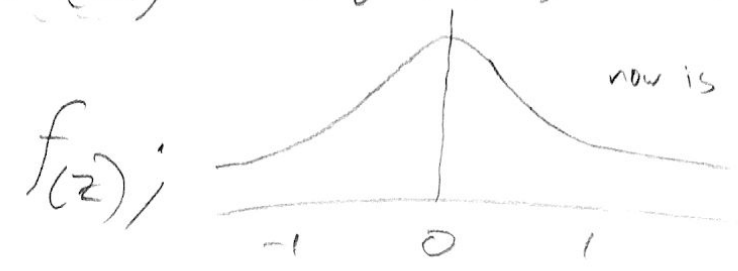
the <u>mixture density</u> is the linear combination of $f_1(z)$ & $F_0(z)$

$$f(z) = \pi_0 f_0(z) + \pi_1 f_1(z) \qquad \pi = \text{probability}$$

the <u>mixture probability distribution</u> for subset Z (the integrated value from distribution of z) linear combination of π_0 & π_1

$$F(z) = \underline{\pi_0 F_0(z)} + \underline{\pi_1 F(Z)}$$

$$F(z) = \pi_0 F_0(z) + \pi_1 F_1(z)$$



* From any subset of $z_1 \rightarrow z_K$, denoted as Z we can calculate the contributions of null or non-null to that probability, as it reflects the linear combinations of π_0 and π_1 at some interval (Z)

($\in$ is set membership, i.e. $A = \{3, 9, 14\}$ then $3 \in A$)

what is the probability that z is null given the subset Z?

$$\Pr\{\text{null} \mid z \in Z\}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

using Bayes theorem:

$$\Pr\{\text{null} \mid z \in Z\} = \frac{\Pr\{z \in Z \mid \text{null}\} \Pr\{\text{null}\}}{P\{z \in Z\}}$$

$$\Pr\{\text{null} \mid z \in Z\} = \frac{\pi_0 F_0(z)}{F(z)}$$

Some finite, calculatable Quantity $F(z)$

$$\phi(z) = Pr\{\text{null} \mid z \in Z\} = \frac{\pi_0 F_0(z)}{F(z)}$$

The ratio of $\pi_0 F_0(z)$, probability of null to the mixture probability distribution $F(z)$ is called the Bayes false discovery rate of z or $\phi(z)$.

If we report some $z \in Z$ as non-null, $\phi(z)$ is the chance that we made a false discovery.

$\phi(z)$ is sometimes written as $Fdr(z)$, stands for false discovery rate.

How is that different than a false positive, or type 1 error?

FP points back to methods, not computational error.

Multiple hypotheses tested! Each probably a false discovery, and some not!

More False-discovery rate

For any z-value we have some...

$P\{\text{null}\} = \pi_0$ calculated — $f_0(z)$ [null]

$P\{\text{non-null}\} = \pi_1$ from z — $f_1(z)$ [non-null]

For any subset Z, we have a density $F(Z)$

null: $$F_0(Z) = \int_Z f_0(z)\,dz$$

non-null: $$F_1(Z) = \int_Z f_1(z)\,dz$$

For this subset, the contributions of null and non-null to any $F(Z)$ is

$$F(Z) = \pi_0 F_0(Z) + \pi_1 F_1(Z)$$

For any observation z in Z, the probability that z is null is

$$\text{Fdr}(Z) = \frac{\pi_0 F_0(Z)}{F(Z)}$$

[While this is very numerical, I think an undergraduate will be able to follow this after a proper statistics course. There are plenty of data sets out there to actually put together a lab for students to do. I have lots of chemical data that could be used.]

Some other ideas for development

Summary of atomic theory and statistics

Harmonic Oscillator Lecture

Lecture on Spectroscopy

Lecture on Heat

[The next step in our understanding is to connect this statistical understanding to atomic statistically theory. There is a way to do this, but I have only completed just a few tiny introductory notes. It stems from atomist theory going back to the mathematicians Boltzmann and later Hilbert. This line of thinking brushes up against the new field of Quantum Bayesianism, see Von Baeyer, Hans Christian, “QBism : the future of quantum physics”. I’m hoping I can write and lecture about this in some intelligible way.]

Here is a little snapshot of an introduction.

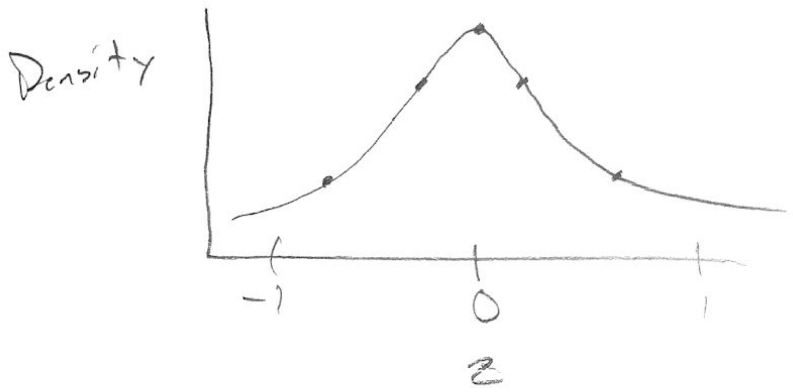
Remember the Born rule!

where v is some volume

$$\int |\psi|^2 dv = 1$$

also The sum of all discrete probabilities equals)

$$\sum P_i = 1 \quad \text{along some probability density}$$



If we model the density, a Bayesian approach, we can solve for any non-continuous interval,

$P(z)dz$ where,

$$\int_z P(z)dz = P_z \quad \left(\text{an individual } P_i\right)$$

$P\{\text{null}\}$ or π_0 as a proportion of the continuous Bayesian density, $F(z)$

$$\pi_0 F_0(z) + \pi_1 F_1(z) = F(z)$$

much like: $\sum P_i = 1$

Does the Born rule apply?

such that $$\int |F(z)|^2 dz = 1$$