Classical Statistics Has Outlived Its Usefulness:

Here's The Fix

William M. Briggs

matt@wmbriggs.com

February 2, 2021

1 Opening Act

Patient walks into the doctor and says, "Doc, I saw that new ad. The one were the people with nice teeth and loose clothing are cavorting. I want to cavort. Ad said, 'Ask your doctor about profitol.' So I'm asking. Is it right for me? Will it clear up my condition?"

The doctor answers: "Well, in a clinical trial it was shown that if you repeated that clinical trial an infinite number of times, each time the same but randomly different, and if each time you calculated a mathematical thing called a z-statistic, which assumes profitol doesn't work, there was a four percent chance that a z-statistic in one of those repetitions would be larger in absolute value than the z-statistic they actually saw. That clear it up for you? Pardon the pun."

Patient: "You sound like you buy into that null hypothesis significance testing stuff. No. What I want to know, and I think it's a reasonable request, is that if I take this drug am I going to get better? What's the chance?"

Doctor: "I see. Let me see if I can clarify that for you. In that trial, it was found a parameter in a probability model related to getting better versus not getting better, a parameter which is not actually the probability but a parameter like it of getting better, had a ninety-five-percent confidence interval of 1.01 to 1.14. So. Shall I write you a prescription?"

Patient: "I must not be speaking English. I'm asking only one thing. What's the chance I get better if I take profitol?"

Doctor: "Ah, I see what you mean now. No. I have idea. We don't do those kind of numbers. I can give you a sensitivity or a specificity if you want one of those."

Patient: "Just give me the pill. My insurance will cover it."

Ladies and gentlemen: the story you have heard is true. Only the names have been changed to protect the guilty.

2 Pleading The Parameter

Ordinary people ask questions like the patient's: Supposing X is true, what is the chance that Y? Answering is often easy. If X = "This die has six different sides, which when thrown must show only one face up", the probability of Y = "a five shows" is 1/6. Casinos make their living doing this.

The professionals who practice statistics are not like ordinary people. They are puzzled when asked simple probability questions. Statisticians really will substitute those mouthfuls about infinite trials or parameters in place of answering probability questions. Then they will rest, figuring they have accomplished something. That these curious alternate answers aren't what anybody wants never seems to bother them.

Here is why this is so.

We have uncertainty about some Y, like the progress of a disease, the topmost

side of a die, the spin of a particle, anything. Ideally we should identify the causes of this Y, or of its absence. If we could know the cause or know of its lack, then the uncertainty we have would disappear. We would know. If the doctor could know all the causes of curing the patient, then he and the patient would know with certainty if the proposed treatment would work, or to what extent.

Absent knowledge of cause there will be uncertainty, our most common state, and we must rely on probability. If we do not know all the causes of the cure of the disease, the best we can say is that if the patient takes the drug he has a certain chance of getting better. We can quantify that chance if we propose a formal probability model. Accepting this model, we can answer probability questions.

We don't provide these answers, though. What we do instead is speak entirely about the innards of the probability model. The model becomes more important than the reality about which the model speaks.

In brief, we first propose a model that "connects" the X and Y probabilistically. This model will usually be parameterized; parameters being the mathematical objects that do the connecting. Statistical analysis focuses almost entirely on those parameters. Instead of speaking of the probability of Y given some X, we instead speak of the probability of objects called statistics when the value of one or more of these parameters take pre-specified values. Or we calculate values of these parameters and act as if these values were the answers we sought.

This is not just confusing, it is wrong, or at least wrong-headed.

Why these substitutions for simple probability questions happen is answered easily. It is because of the belief that probability exists. Probability, some say, exists in the same way an electric charge exists, or in the way the length of

3

the dollar bill exists. Observations have or are "drawn from" "true" probability distributions. If that is true, and probability really does exist, then the parameters in those parameter models also exist, or are measures of real things. This being so, it makes sense to speak of these real objects and to study them, as we might, say, study the chemical reactions that make flagellum lash.

The opposite view is that probability does not exist, that it is entirely epistemological, a measure of uncertainty. Probability is a (possibly quantified) summary of the uncertainty we entertain about some Y given some evidence X. In that case, it does not make sense to speak of model parameters, except in the formal model building steps, steps we can leave to the mathematicians.

These two beliefs, probability is real or in the mind, have two rough camps of followers. The one that believes probability exists flies the flag of Frequentism. The one that says it doesn't flies the flag of Bayes. Yet most Bayesians, as they call themselves, are really frequentist sympathizers. When the data hits the code, the courage of their convictions withers and they cross to the other side and become closet frequentists. Which is fair enough, because frequentists do the same thing in reverse when discussing uncertainty in parameters. The result is a muddle.

Let me first try to convince you probability doesn't exist. Then I'll explain the two largest generators of over-certainty that come from the belief in probability existing. Finally, I'll offer the simple solution, a solution which has already been discovered and is in wide-spread use, but not by statisticians or those who use statistical models.

3   Roll With It

You do not have a probability of being struck by lightning. Nobody does. There is no probability an electron will pass through the top slit of a two-slit

experiment. There is no chance for snake eyes at the craps table.

There is no probability a random mutation will occur on a string of DNA and turn the progeny of one species into a new species. There is no probability a wave function for some quantum system will collapse into a definite value. There isn't any chance you have cancer. There isn't any chance that physical (so-called) constants, such as the speed of light, took the values they do so that the universe could evolve to be observed by creatures like us.

If probability existed in an ontological sense, then things would have probabilities. If a thing *had* probability, like you being struck by lightning, then probability would be an objective property of the thing, like a man's height or an electron's charge. In principle, this property could be measured, given specified circumstances, just like height or charge. Probability would have to be more than just a property, though. It either must act as a cause, or it must modify causes to the thing of interest. It would, for example have to draw lightning toward you in some circumstances and repel it in others, or it would have to modify the causes that did those things. If probability is a direct cause, it has powers, and powers can be measured, at least in principle. If probability only modifies causes, it can either be adjusted in some way, i.e. it is variable, or it is fixed. In either case, it should be easy, at least for simple systems, to identify these properties.

If things *have* probability, what part of you, or you plus atmospheric electricity, or you plus whatever, *has* the probability of being struck by lightning? The whole of you, or a specific organ? If an organ, then probability would have to be at least partly biological, or it would be able to modify biology. Is it adjustable, this probability, and tunable like a radio so that you can increase or decrease its strength? Does some external cause act on this struck-by-lightning probability so that it vanishes when you walk indoors? Some hitherto hidden force would

have to be responsible for this. What are the powers of this cause and by what force or forces does it operate? Is this struck-by-lightning probability stored in a different part of your body than the probabilities of cancer or of being audited by the IRS? Since there are many different things that could happen to you, each with a chance of happening, we must be swarming with probabilities. How is it that nobody has ever seen one?

Here is a statement: "There are four winged frogs with magical powers in a room, one of whom is named Bob; one winged frog will walk out the door." Given this statement, what is the probability that "Bob walks out"? If probability is in things, how is it *in* non-existent winged frogs? Some say that Germany would have won World War II if Hitler did not invade Russia. What is the probability this is true? If probability exists, then how could probability be *in* a thing that never happened?

I was as I wrote this either drinking a tot of whiskey or I was not. What is the probability I was drinking that whiskey? Where does the probability live in this case, if it is real: in you, in me, in the whiskey? There is an additional problem. Since I know what I was doing, the probability for me is extreme, i.e. either 0 or 1, depending on the facts. The probability won't be either number for you since you can't be certain. Probability is different for both of us for the same event. And it would seem it should be different for everybody who cared to consider the question. Probability if it exists must be on a continuum of a sort, or perhaps exist as something altogether different. Yet since probability can be extreme, as it is for me in this case and is for you, too, once you learn the facts (I was not drinking), it must be, if probability is real, that the probability just "collapsed" for you. Or does it go out of existence?

Well, maybe probability doesn't exist for any of these things, but it surely must exist for quantum mechanical objects, because, as everybody knows, we

calculate the probability of QM events using functions of wave functions (functions of functions!), and everybody believes wave functions are ontologically real. Yet we also calculate probabilities of dice rolls as functions of the physical properties of dice, and probability isn't in these properties, because if we're careful we can control outcomes of dice throws. We can know and manipulate all the causes of dice throws. We know we can't with QM objects.

Probability in QM isn't the wave function, it's a function of the wave function, and also of the circumstances of the measurement that was made (the experiment). The reason we think QM events have probability is that we cannot manipulate the circumstances of the measurement to produce with certainty stated events, like we can with dice (and many things), by carefully controlling the spin and force with which dice are thrown. Again, with dice, we can control the *cause* of the event, with QM we cannot. The results in QM are always uncertain; the results with dice need not be. Since Bell, we know we cannot know or control all the causes of QM events (the totality of causes). This has caused some people to say the cause of QM events doesn't exist, yet things still happen, therefore that this non-existent cause is probability. Some will make this sound more physical by calling this strange causal-non-causal probability *propensity*, but given all the concerns noted above, it is easy to argue propensity is just probability by another name.

Whether or not that is true, and even if these brief arguments are not to convince you probability does not exist, and accepting philosophers constantly bicker over the details, I am hoping it is clear that if in any situation we did know the cause of an event, then we would not need probability. Or, rather, conditional on this causal knowledge, probability would always be extreme. At the least, probability is related to the amount of ignorance we have about cause. The stronger the knowledge of cause, the closer to extreme the probability is. In

any case, it is knowledge of cause which is of the greatest importance. Searching for this knowledge is, after all, the purpose of science.

The main alternate view of probability is to suppose it is always a statement of evidence, that it always epistemological. Probability is about our uncertainty in things, and not about things as they are in themselves. Probability is a branch of epistemology and not ontology. Bruno de Finetti famously shouted this view, and after an English translation of his rebel yell appeared in 1974, there was an explosion of interest in Bayesian statistics, the theory which supposedly adopts this position.[1] Everybody quotes this, for good reason (ellipsis original):

> The abandonment of superstitious beliefs about the existence of the Phlogiston,the Cosmic Ether, Absolute Space and Time,...or Fairies and Witches was an essential step along the road to scientific thinking. Probability, too, if regarded as something endowed with some kind of objective existence, is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs.

There were others beside de Finetti, like the physicist E.T. Jaynes, economist John Maynard Keynes, and the philosopher David Stove, who all held that probability is purely epistemological.[2] Probability is an objective classification or quantification of uncertainty in any proposition, conditional only on stated evidence.

After de Finetti's and these others' works appeared, and for other reasons, many were ready to set aside or give less oxygen to frequentism, the practice of statistics which assumes that probability is real and only knowable "in the limit". Room was made for something called subjective Bayesianism. Bayesianism is the idea probability is epistemic and that it can be known subjectively. Probability is therefore mind dependent. Yet if probability is wholly subjective,

a bad meal may change the probability of a problem, so we have to be careful to define subjectivity.

Frequentism, with its belief in the existence of probabilities, is far from dead. It is the form of and practice of statistics taught and used almost everywhere. All Bayesians start as frequentists, which might be why they never let themselves break entirely free from it.

Now the philosophical position one adopts about probability has tremendous consequences. You will have heard of the replication crisis afflicting fields which rely heavily on statistics, whereby many results once thought to be novel or marvelous are now questioned or are being abandoned.[3] Effects which were once thought to be astonishing shrink in size the closer they are examined. The crisis exists in part because of the belief probability is real. Even beside this crisis, there is massive over-certainty generated in how statistics is practiced.

4 What Probability Is And Isn't

All probability can be written in this schema:

$$\Pr(Y|X),$$

where Y is the proposition of interest, and X *all* the information that is known, assumed, observed, true, or imagined to be true, information that is thought to be probative of Y. Included in X—and here is what is most forgotten in the heat of calculation—are the rules of grammar, definitions of words and symbols, mathematical or in some other language, and all other tacit knowledge about X and Y which is thought too obvious to write down. X must always be present. However useful as shorthand, it is an error to write, *sans* X, $\Pr(Y)$, notation which suggests that probability exists and that Y has a single unique probability

that can be discovered.

All probability is conditional on some X. If this is doubted, it is an interesting exercise to try and write a probability of some Y without an X. Pick any proposition Y and try to show it has a unique probability without any conditions whatsoever; i.e. that $\Pr(Y)$ for your Y exists. You will soon discover this cannot be done. Recall that all tacit premises of grammar and definitions about Y must be included as premises X. Try Y = "This proposition is false." What is $\Pr(Y)$? This is hours of fun for the kids.

What is the probability of Y = "I have cancer"? The question is pressing and relevant. It is clear that your knowledge of whether you have cancer is not the same as whether you actually have cancer. Since the question is of enormous interest, we want to give an answer. Information that is relevant to the causes of cancer present themselves: "I'm a man, over fifty; I smoke and am too fond of donuts..." This amorphous list is comprised of what you have heard, true or not, about causes of cancers. You might reason

$$\Pr(\text{Cancer}|\text{I Smoke}) = \text{ good chance.}$$

There are no numbers assigned. No numbers can be assigned, either; none deduced, that is. To do that we need to firm up the X to create a mathematical tie between it and Y.

The real interest in any probability calculation is therefore in X: which X count for *this* Y. Ideally, we want to know the cause, the reason, for Y's truth or its falsity. Once we know a thing's cause or reason for existence, we are done. Barring this perfect state of knowledge, we'd like to get as close as we can to that perfection. Science is the search for the X factor.

The choice of Y is free. This part of probability can be called subjective. Once a Y is decided upon, the search for the X that is Y's cause, or is in

some other way probative, begins. Probability can be called subjective at this step, too, for what counts as evidence can itself be uncertain. Sports are a good example. I think it's likely the Tigers will win tomorrow, while you say it isn't. We list and compare all our premises X, some of which we both accept, some we don't and which are individual to each of us. If I agree with *all* your evidence, then I must agree with your probability. In this way, probability is not subjective. Probability is not decision, or rather it comes before decision, so that even if we agree in *all* X, and therefore our probabilities match, we might differ in what decisions we make conditional on this.

If we agree that X = "This is a 6-sided...etc." and that Y = "A five spot shows" then it would be absurd if you said the probability was, say, 4/6 while I said 1/6. If you did say 4/6 it must be because you have different, tacit, X than I have. Our premises do not agree. But if they did agree, then probability is no more subjective in its calculation than is algebra once the equation is set. Once the X and Y are fixed, the answers are deduced.

One form of subjective probability asks a man to probe his inner feelings, which become his X, to use in calculating the probability of Y. The process is called "probability elicitation", which makes it sound suitably scientific. And it can be, though it can lead to forced quantification, hence over-certainty. Since the choice of X is free, there is no problem *per se* with this practice, except that it tends to hide the X, which are the main point of any probability exercise. Subjective probability becomes strange when some require the mind to enter into the process of measurement, as some do with quantum mechanics.[4] That subject is too large for us today.

In practice, there is little abuse of subjective probability in ordinary statistical problems. Mostly because there is nothing special about Bayesian probability calculus itself. It is just probability. Bayes is a useful formula for computing

probabilities in certain situations, and that's it. Bayes is supposed to "update" belief, and it can, but the formula is just a mechanism. We start with some X_1, probative about Y. We later learn X_2, and now we want $\Pr(Y|X_1X_2)$. The Bayes formula itself isn't strictly needed (though no one is arguing for discarding it) to get that. We always want $\Pr(Y|X)$ whatever the X is, and whenever we get it. If we call X the "updated information", the union of X_1 and X_2, then it may be that Bayes formula provides a shortcut to the calculation, and again it may not.

The real departure of Bayes from frequentism is not in the questions of the subjectivity of probability, for the same subjective choices must be made by frequentists in picking their Y and X. It's that Bayes insists all uncertain propositions must have conditional probability in the epistemic sense, whereas in frequentism things that were said to be "fixed" do not, and are in fact forbidden by the precepts of the theory to have anything to do with probability. The real cleft is whether or not the uncertainty in parameters of probability models should be quantified with probability. Bayesians say yes, frequentists no. Just what are these parameters?

5 Modeling Agency

A model or theory is a list of premises X which are probative of Y. If a number and not just a word for the probability is desired, some way to connect the X to the Y such that quantities can be derived must be present. This can be trivial counting as when

$$X = \text{"This machine must take 1 of } n \text{ states"}$$

to

$$\Pr(Y = \text{``This machine is in state } j\text{''}|X) = 1/n.$$

The deduction to $1/n$ can be done by calling to the symmetry of logical constants[5]. There is no looseness in X: it is specific. This is important: we must take the words as they are without addition (except of course their definitions). That is, there is no call to say "Well, some machines break", which may be true for some machines, but it is not supposed here. Probability is always, or should always, be calculated on the exact X specified, and nothing else.

More complex probability models use parameterized distributions, an ubiquitous example being the normal, the familiar bell-shaped curve. It is properly said to represent uncertainty in some observable Y. But often people will say Y *is* normal, as if the observable *has* the properties of a normal, which is another way of saying probability exists. If the probability exists, the parameters of normal distribution must also exist, and must in some way be part of the observable, or the observable plus measurement, as suggested above. The manner in which this might be true is, of course, never really specified. It's simple enough to show that this can't be true, or at least that it can't be known to be true.

Our X in this case might be "The uncertainty in Y is represented by a normal with parameters 10 and 5". Any Y will do, such as Y = "The measurement is 7." We can calculate $\Pr(Y|X)$, which is equal to 0. And is equal to 0 for any singular or point measurement. Which is to say, given X, the probability of any point measurement is 0. This happens because, as is well known, the normal gives all its probability to the continuum and none to actual measurements. There is no difficulty in the math, but this situation presents a twist to understanding probability. Any measurement we might take is finite and discrete; no instrument can probe reality to an infinite level, nor can we store

infinite information. There is therefore no set of actual measurements that can ever conclusively prove an observable *is*, or *is from*, a normal distribution.

The counter to this is to appeal to the central limit theorem and say collections of measurements of observables are comprised, or are made by, many small causes. In the limit, it is therefore the case the observable *is* a normal. This proof is circular. There is no justification for assigning the normal distribution in the first place to an observable because we don't know it will always and forevermore be created by these small additive causes. The only time we can know we have a normal is when we are at that happy limit where we know all there is to know about the observable. At which point we will no longer need probability. Beside all that, there is no proof, and every argument against, anything lasting forever. This real-world finiteness is granted, but it still claimed y is normal, with the feeling—it is no more than that—that the normal is what in part generates or causes y. This is a strange view is never really fleshed out.

These same objections about finiteness apply to any probability model of measured observables that are not deduced and which are merely applied *ad hoc*. Most models are in fact *ad hoc*. Of course, the normal, and many similar models, are often employed in situations where the measurements are known to be finite and discrete. These models can be and are useful, as long as we are happy with speaking of probabilities of intervals and not singular points, and we're aware at all times the models are approximations.

Suppose in fact we are willing to characterize our uncertainty in some observable y with a normal distribution, the familiar bell-shaped curve, with parameters 0 and 1 (the lower case y is shorthand to things like Y = "$y > 0$"). These parameters specify the center of the bell and give its spread. These suppositions about the model and parameters are our X. It's then easy to calculate things like $\Pr(y > 0|X) = 0.5$. It's also clear that we have perfect certainty in

the parameters: they were given to us. It would be an obvious error to say that the uncertainty we have in these parameters, which is none, is the same as the uncertainty we have in the observable, which is something.

Yet this mistake is made in practice. To see how let's expand our model. We have the same observable and also a normal, only this time we don't know what values the parameters (μ, σ) take. We want to know $\Pr(y \in s | \mathrm{X})$ for some set s, where X is the assumption of the normal. Since the parameterμ can take any value on the real line, and σ any value on the non-negative part of the real line, there is no way to calculate this probability of $y \in s$. Something first has to be said about the parameters. Above we dictated them, which is a form of Bayesianism. They may have even been deduced from other premises, such as symmetry in some applications. That deduction, too, is Bayes. These additional premises fall into X, and the calculation proceeds.

Any premises relevant to the parameters can be used. When these premises put probabilities on the parameters the premises are called "priors"; i.e. what we know or assume about the parameters before any other information is included. A common set of prior premises is to suppose that $\mu \sim \mathrm{N}(\nu, \tau)$, another normal distribution where the "hyper-parameters" (ν, τ) are assumed known by fiat, and that $\sigma \sim \mathrm{IG}(\alpha, \beta)$, an inverse gamma (the form is not important to us), and again where the hyper-parameters (α, β) are known (or specified).

A frequentist does not brook with any of this, insisting that once the probability model for y is specified, the parameters (μ, σ) come into existence, or they always existed, only we just now became aware of them (though not of their value). These parameters must exist since probability exists. The parameters have "true" values, and it is a form of mathematical blasphemy to assign probability to their uncertainty. The frequentist is then stuck. If he isn't handed the true values by some oracle, he cannot say anything about $\Pr(y \in s | \mathrm{X})$, where

for him X is only evidence that he uses the normal.

The Bayesian can calculate $\Pr(y \in s | \mathrm{X}_b)$, the subscript denoting the different evidence than that assumed by the frequentist, X_f. The values of (ν, τ) and (α, β) are first spoken, which gives the probabilities of the parameters. The uncertainty in these parameters is then integrated out using Bayes's formula, which produces $\Pr(y \in s | \mathrm{X}_b)$, which in this case has the form of a t-distribution, the parameters of which are functions of the hyper-parameters. The math is fun, but beside the point.

The frequentist objects that if the priors were changed, the probability of $y \in s$ will (likely) change. This seems a terrible and definitive objection to him. The criticism amounts to this, in symbolic form: $\Pr(y \in s | \mathrm{X}) \neq \Pr(y \in s | \mathrm{W})$, where $\mathrm{X} \neq \mathrm{W}$. This objection pushes on an open door. Since probability changes when the probative premises change, *of course* the probability changes when the priors change. But to the frequentist, probability exists and has *true values*. These priors might not give the true values, since they are arbitrary. Even granting that objection, the frequentist forgets the normal model in the first place was also arbitrary, a choice between hundreds of other models. It too might not give the true probability.

We'd be stuck here, except that the frequentist allows that previous observations of y are able to give some kind of information about the parameters. The Bayesian says so too, but the kind of information for him is different.

The frequentist will use previous observations to calculate an "estimate" of the parameters. In the normal model, for the μ it is usually the mean of the previous y; for the σ it is usually the standard deviation of those y.[6] The word *estimate* implies a true value exists, because probability exists. The frequentist admits there is uncertainty in the guesses, and constructs a "confidence interval" around each guess. Here is the definition of a 95% confidence interval: if you

repeated the experiment, or suite of measurements, or the set of circumstances that gave rise to the observed y, an infinite number of times, each time exactly the same but randomly different, and each time calculating the estimate and the confidence interval for the estimate, then in that infinite set of confidence intervals 95% of them will "cover" the true value of the parameter.

What of *this* confidence interval? The *only* thing that can be said is that either the true value of the parameter is in it, or it isn't. Which is a tautology and always true, and therefore useless.

No frequentist ever in practice uses the official definition of the confidence interval, proving that no frequentist has any real confidence in frequentist theory. Every frequentist instead interprets the confidence interval as a Bayesian would, as giving the chance the true value of the parameter is in this interval. The Bayesian calculates his interval, called a "credible interval', in a slightly different way than the frequentist, using the priors and Bayes theorem. In the end, and in many homely problems, the intervals of the frequentist and Bayesian are the same, or close to the same. Even when the intervals are not close, there is a well known proof that shows that as the number of the observations increases, the effects of the prior on the interval vanish.

So, given these at least rough agreements, what's the point of mentioning these philosophical quibbles which excite statisticians but have probably bored the reader?

There are two excellent reasons to bend your ear. The first is that, just as frequentists became occult Bayesians in interpreting their results, the *Bayesians became cryptic frequentists when interpreting theirs!* That the Bayesians also speak of a "true" value of their parameter also means they don't take their theory seriously, either.

Even this wouldn't be a problem, except for a glaring omission that seems

to have escape everybody's attention. This is the second reason to pay attention. We started by asking for $\Pr(y \in s | \mathrm{X})$. The frequentist supplied X_f and the Bayesian X_b. Both calculated intervals around estimates of μ. Both then stopped. What $\Pr(y \in s | \mathrm{X}_f)$ or $\Pr(y \in s | \mathrm{X}_b)$ is we never learn. The parameters have become everything. Actually, only one parameter: the second, σ, is forgotten entirely.

6 Testing Our Patience

This parameter-centric focus of both frequentists and Bayesians has led to many difficulties. The first is "testing".

The patient we met at the beginning had Y = "I am cured" and X = "I take profitol" and wanted $\Pr(\mathrm{Y}|\mathrm{X})$. The doctor instead first told him the results of a statistical test. The simplest such test works like this. Cures and failures when taking profitol or a placebo happen. Since we don't know all the causes of cures or failures, it is uncertain whether taking the drug will cause a cure. The uncertainty in the cause is quantified with a probability model, or in this case two probability models. One has a parameter related to the probability of a cure for the drug, and the second a parameter related to the probability of a cure for the placebo. These parameters are often called *the* probabilities of a cure, but *they are not*; if they were, we would know $\Pr(\mathrm{Cure}|\mathrm{Drug})$ and $\Pr(\mathrm{Cure}|\mathrm{Placebo})$, and we'd pick whichever is higher.

The test begins with the notion that the probabilities are unknown and must be estimated. But we never want to *estimate* the probability (except in a numerical approximation sense): we want $\Pr(\mathrm{Cure}|\mathrm{X})$ period, where X is everything we are assuming. X includes past observations, the model assumptions, and which pill is being swallowed. The problem here is the overloading of the word *probability*: in the test it stands for a parameter, and it also stands for the

18

actual conditional probability of a cure. Confusion arises through this double meaning.

In other words, what we should be doing is calculating Pr(Cure|Drug(X)) and Pr(Cure|Placebo(X)). But we do not.

Instead we calculate a statistic, which is a function of the estimates of the two parameters. There are many possible non-unique choices of this statistic, with each giving a different answer to the test. One statistic is the z-statistic. To calculate its probability, it is *assumed* the two parameters are equal. Not just here in the past observations, but everywhere, for all possible observations. If probability exists, these parameters exist, and if they exist they might be equal. Indeed, they are said to be equal. With these assumptions, the probability of seeing a z-statistic larger in absolute value than the one we actually saw is calculated. This is the p-value.[7]

If the p-value is smaller than the magic number, which everybody knows and which I do not have to repeat, it is *decided* that the two parameters representing probability of cure *are* different, and by extension, the probability of cures *are* different.

The Bayesian sees one weakness with this: the test puts things backwards. It begins by assuming what we want to know, and uses some odd decision process to confirm or disconfirm the assumption. We do not know the probability the two parameters are unequal, or that one is higher than another, say. The Bayesian might instead calculate $\Pr(\theta_d > \theta_p|X)$ (the notation being obvious). This doesn't have to be a strict greater-than, and can be any function of the parameters that fits in with whatever decisions are to be made. For instance, sometimes instead of this probability, something called a Bayes factor is calculated. The idea of expressing uncertainty in the parameters with probability is the same.

The innovation of these parameter posteriors (for that is their name) over testing is two-fold. First, it does not make a one-size-fits-all decision like p-values and declare with finality that parameters are or aren't different, or, in the bizarre falsification language of testing, that it hasn't been disproved they are the same. Second, it puts a quantification on the probability on a potential question of interest; i.e. whether the parameters really are different.

The Bayesian has stopped short: these posteriors do not answer *the* question of interest. We wanted the probabilities of cures, not whether some dumb parameters were different. Why not just forget all this testing and focus on parameters and calculate these probabilities?

Alas, no. Instead of testing, what might be calculated is something called a risk ratio, or perhaps an odds ratio. The true "risk" of a cure here is $\Pr(\text{Cure}|\text{Drug(X)})/\Pr(\text{Cure}|\text{Placebo(X)})$. This is fine one-number summary of the two probabilities, albeit with a small loss of information. The model-based risk ratio is not this, and is instead a ratio of the parameters, which again are not the probability but which are called probabilities. Since the probabilities are never calculated, and instead estimates of the parameters are, an estimate of the model risk ratio is given, along with its confidence or credible interval. The big but is that this is an interval of the ratio of parameters, which exaggerates certainty. Since we can, if we wanted to, calculate the ratio of the probabilities themselves, it isn't even needed.

This simple example is multiplied indefinitely because almost all statistical practice revolves around parameter-centric testing, or parameter estimation. Parameters are not the problem, though, because they are necessary in models. Since at least because they can be never observed, and since they don't answer probability questions about the observable, they should not be the primary focus.

It is parameters or functions of parameters which are reported in almost all analyses, it is the parameters which are fed into decisions, including formal decision analysis; it is even in many cases the parameters which become predictions, and not observables. All this causes massive over-certainty, and even many errors, mainly about ascribing cause to observations.

Here is a simple example of that over-certainty, using regression, that ubiquitous tool. Regression assigns a parameter β to a supposed or suspected cause, such a sex in a model of income. The parameter in this case will represent the *difference* in sexes. The regression will first test whether this parameter is 0. If it is decided, via a large p-value, that this parameter is 0, then it will be announced "There is no difference in incomes between males and females." If the p-value is instead wee, then it will be said "Males and females have different incomes." The face over-certainty is obvious.

Next comes the point estimate of the parameter. Suppose income is measured in thousands, and that the estimate of the parameter is 9. It will be announced as definitive that "Males make on average \$9,000 more than females." The confidence interval is, let's say, $(8, 10)$. It will be announced "There is a 95% chance males make from between \$8,000 to \$10,000 more than females." Even ignoring the misinterpretation of the confidence interval, this is still wrong.

These numbers are about the parameter, and not income. The predictive probability $\Pr(M > F | X)$ will not be equal to $\Pr(\beta > 0 | X)$. It depends on the problem, but experience shows the latter probability is always much larger than the former. The probability of the $\beta > 0$ may be close to extreme, while the probability of the observables, M>F incomes, may be near 0.5, a number which expresses ignorance about which sex makes more. Again, certainty in the parameters does *not* translate into certainty in the observables. Worse, the 95% predictive interval in income differences will necessarily be wider than the

interval of the parameter. Experience shows that for many real-life data sets, the observable predictive interval is 4-10 times wider than the parametric interval. How much over-certainty really exists in published literature has not yet been studied, but there is no sense that it is small. In the example we stared with, with the normal $(0,1)$ model, the predictive interval is infinitely larger than the parametric, which is 0.

This same critique can be applied to *any* probability model that is cast in its parametric and not predictive, i.e. probability, form. The reason for the parameter preference is because of the belief probability exists.

Observations, it is said, are "drawn from" probability distributions, which are a feature of Nature. If we knew the true probability distribution for some observable, then we'd make optimal decisions. Again, if probability exists, parameters exist, and it is a useful shorthand to speak of parameters and save ourselves the difficulty of speaking of probabilities, which would be equivalent in a causal sense. That β in the regression example is taken to be proving causes exist that make males earn more than females—which might be true, but it is not proven. Any number of things might have caused the income differences.

If we understood what was causing each Y, then we would know the true state of nature. There is in statistical practice a sort of vague notion of cause. In the drug example, if the test is passed, some will say that the drug is better at causing cures than the placebo. Which, of course, might be true. But it cannot be proven using probability.

In the set of observations we are imagining some cures were caused by the placebo; now whether this was the placebo itself or that the placebo is a proxy for any number of other causes is unimportant. The drug group we can assume saw proportionately more cures. How many of those cures in that group were caused by the placebo? All of them? None? There is no way to know, looking

only at the numbers.

If we look outside the numbers to other evidence, we might know whether the drug was a cure. Or sometimes a cure, since it will usually be the case the drug does not cure all patients. We consider our knowledge of chemistry, biology, and other causal knowledge. If all that tacitly becomes part of X, then we can deduce that some of the cures in the drug group were caused by the drug. But then it becomes a question of why everybody wasn't cured, if the drug is a cause of cures. It must be that there are some other conditions, as yet identified or not assumed in X, that are different across individuals. In effect, the discussion becomes of what is blocking the drug's causal powers.

There is a prime distinction, well known, between observations that were part of a controlled environment and those which were merely observed.[8] Physicists understand control well. In measuring an effect in an experiment, every possible thing that is known or assumed to cause a change in the observable is controlled. Assuming all possible causes have been identified, then the cause in this experiment may be deduced. Of course, if this assumption is wrong or ignored, then it is always the case that something exterior to our knowledge was the true cause. If if it right but ignored, then who can disprove that inter-dimensional Martian string radiation, or whatever, wasn't the real cause? It is thus always possible something other than the assumed cause was the true cause of any observation. It is also the case that this complete openness to external causes is silly.

We end where we began. If we knew the causes of the observable Y, we do not need probability. If we do not know all the causes of Y, we are uncertain, and thus need probability. Parameter-based testing and parameter estimation are not probability of observables, but strange substitutes which cause over-certainty.

7 The Fix Is In

The fix is simplicity itself. Instead of testing or estimating, calculate $\Pr(Y|X)$.[9] Give the probability of a cure when taking the drug; express the probability males make more than females with a probability. Every statistical model can be cast into this predictive approach. In Bayesian statistics it is called calculating the predictive posterior distribution. Some do this, but usually only when the situation seems naturally like a forecasting problem, like in econometrics. It works for *every* model.

Even if you haven't been convinced by all the earlier examples for the excellency of this suggestion, think of this. When we are analyzing a set of old observations, *we know all about those observations.* Testing and estimation are meant to say something about the hidden properties of these observations. If these past observations are the only measurements we will ever take, then we do not need testing or estimation! If we observed males had higher mean income than females, then we are 100% certain of this (measurement error can be accounted for if it arises). It is only because we are uncertain of the values of observations not yet made known to us that we bothered with the model in the first place. That demands the predictive, or probability, approach.

Computer scientists have long been on board with this solution. Just ask them about their latest genetic neural net deep big learning artificial intelligence algorithm. (Computer scientists are zealous in the opposite direction of statisticians.) These models are intensely observable-focused. Those scientists who must expose their theories to reality on a regular basis, like meteorologists, are also in on the secret. The reason meteorologists' probability predictions improve, and why the models of say sociologists do not, is because meteorologists test their models against real life on a daily basis.

I pick on sociology because they are heavy users of statistical models. They

24

will release a model after it passes a test, which if you read the discussion sections of their papers means to them that the theory they have just proposed is true. Nobody can easily check that theory, though, since it is cast in the arcane statistical language of testing or estimation. Anybody can check if the weather forecast is right or useful. If instead the sociologist said, "If you do X, the probability of Y is 0.9," then anybody with the capability of doing X can check for themselves how good or useful the model really is. You don't need access to the original data, either, nor anything else used in constructing the model. You just need to do X.

The transparency of casting models in terms of probabilities, i.e. in their predictive form, may be one reason why this practice hasn't been adopted. One can be mighty and bold in theorizing, but when one is forced to bet, well, the virtue of humility is suddenly recalled.

Incidentally, if you have to ask your doctor whether an advertised pill is right for you, you might want to consider finding a more knowledgeable doctor.

Notes

1. The book is Bruno de Finetti, 1974. *Theory of Probability*, (translation by A Machi and AFM Smith of 1970 book), Volume 1, Wiley, New York; quote from p. x.

2. Necessary reading are Jaynes's 2003 *Probability: The Logic of Science*, Cambridge University Press, Jim Franklin's 2001 "Resurrecting logical probability", *Erkenntnis*, Volume 55, Issue 2, pp 277–305, and John Maynard Keynes's 2004 *A Treatise on Probability*, Dover Publications. Stove is the dark horse here, and I could only wish his *The Rationality of Induction* were better known.

3. There are any number of papers on the replication crisis. A typical one is Camerer et al., Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2 (9), pp. 637–644.

4. Christopher A. Fuchs is among those trying to marry subjective Bayesian probability with

quantum mechanics. See Caves, C.M., C.A. Fuchs, and R. Schack, 2001. Quantum probabilities as Bayesian probabilities, DOI: 10.1103/PhysRevA.65.022305.

5. David C. Stove, 1986, *The Rationality of Induction*, Oxford University Press. The second half of the book is a brilliant justification of probability as logic which gives this rare proof. Note this doesn't have to be a real machine, so we don't need notions of symmetry; rather, symmetry is deduced from the premises.

6. So common are these estimates, that it has become customary to call the parameters *the* mean and standard deviation; however this is strictly a mistake.

7. I have a collection of anti-p-value arguments in "Everything Wrong With P-Values Under One Roof", 2019, In *Beyond Traditional Probabilistic Methods in Economics*, V Kreinovich, NN Thach, ND Trung, DV Thanh (eds.), Springer, pp 22—44. The use of p-values is changing. Even so staid an organization as the American Associations of Statisticians has begun issuing warnings that probability-as-real measures like p-values should not be relied upon. See Wasserstein, R.L. & Nicole A. Lazar, 2016. The ASA's statement on p-values: context, process,and purpose. *The American Statistician*, DOI: 10.1080/00031305.2016.1154108. Not to be missed is the 2019 *Nature* article "Scientists rise up against statistical significance" by Valentin Amrhein, Sander Greenland, Blake McShane, which relates how over 800 scientists (I am one of them) signed a statement asking for the retirement of the phrase "statistically significant".

8. Some have embraced the notion that cause should be paramount in statistical analysis; a notable example is Judea Pearl in his *Causality*, 2000, Cambridge University Press. These changes in focus make an excellent start, and if there is any problem it is that the existence of probability is still taken for granted.

9. Which X to include is a decision, like any decision, and should be treated as such. X added that do not change the probability of Y, or do not change it such that any decision you would make using the probability would not change, would not be included.

Briggs is the author of Uncertainy. *Get some. Thanks to Philip Pilkington and David Berlinkski for providing comments to an earlier draft.*