

UNRESOLVED ISSUES OF FUNDAMENTAL STATISTICAL INFERENCE

ZULFIKAR MOINUDDIN AHMED

You're sayin' those words like you
hate me now (wo-oah) Our house is
burning when you're raisin' hell
(wo-oah) Here in the ashes your
soul cries out (a-a-ah) But don't be
afraid of these thunderclouds These
thunderclouds, oh, no These
thunderclouds, oh, no, no

LSD, "Thunderclouds"

1. WHAT IS STATISTICAL INFERENCE?

Suppose you have your theory A of all things of Heaven and Earth, and you love this theory. Your conviction level for the theory is enormous. You think that this theory of yours, A will give such deep insight into the nature of existence that the entire human race will be stunned and grow wings and fly away into another world where peace and harmony reigns for millions of years and will end all difficulties of history. So you are pretty enthusiastic about your theory A .

Now suppose your theory is quite complex in its details but among the important measurable concepts of the theory is a variable v . I don't know the details of your theory A , so I won't try to explain what role this v plays to explain all things of Heaven and Earth.

Now suppose you obtain grants from the governments of Earth who are all enthusiastic about your theory and you obtain the services of many lab scientists from Europe, Americas, Africa, Asia, Oceania. They particularly are interested in the variable v because they think you are right. You're quite fortunate, aren't you? My theories never get so much enthusiasm.

The problem of statistical inference for v is to determine how to get Nature's true value of v from N measurements v_1, \dots, v_N . In other words your theory is some intellectual scheme that man can understand. The enthusiasm of all these lab scientists suggests that they respect your theory as a candidate for how Man finally understands all things of Heaven and Earth with Man's concepts. There is wild enthusiasm for the idea that the actual universe as A posits, and so Nature herself contains an intrinsic v and the problem of statistical inference is to gain confident value for v from a finite set of measurements v_1, \dots, v_N .

2. IMPORTANT DIGRESSION

I am not actually a Rationalist zealot. I strongly do not agree that Reason is part of Human Nature. Nor do I believe that scientific theories are substitutes for faith in mythological subjective worlds. The latter are not 'pre-scientific' but human. Science cannot substitute for these. Science is powerful on one aspect of human experience, one limited view of the universe, where the quantities measured have qualities such as repeatability. Human beings are the sorts of beings who cannot actually live in the Scientific world and require a far more complete subjective reality that have many non-repeatable phenomena. But I am not bringing out these issues for the first time. Friedrich Nietzsche's first youthful work, *The Birth of Tragedy from the Spirit of Music* addresses the critique of science from an *artistic viewpoint* masterfully.

At the same time, Science is most certainly not trivial pursuit, and scientific worldview is part of human civilisation. It is necessary and delightful, but will never be complete for a healthy human soul.

In particular, I am extremely skeptical that there was a permanent transformation during the Scientific Revolution begun by Rene Descartes. I expect mythological worldview to reclaim humanity periodically for that is much more comfortable for human souls.

This interplay between mythological worldviews and scientific worldviews will surely find convergence in the long run of thousands of years.

3. THE FAITH OF SCIENTISM REGARDING LIFE BEGINS AT CONCEPTION IS WRONG

If you lived in a desert village far from the ocean, and had never seen the ocean, and concluded quite wisely – in your estimate – that the ocean is a figment of people's imagination would be simply wrong. When I was very young, I was in tropical warm Bengal and thought that people who claimed to have winters colder than the inside of the freezer were bamboozling me with fibs. I faced a cold winter in Brooklyn in 1987 or 1988 and that showed me the error of my previous beliefs.

I have faith that my soul existed before my birth to my mother and father, and I have faith about experiences of my soul before birth. Those people who feel that they are sure that this is fiction are deluded for *their faith* that life begins at conception is simply wrong. They will say they need evidence before they believe otherwise. That's their problem. When they have more evidence they will know better. I do not need evidence for things that are obvious and self-evident to me.

There might seem a certain satisfactoriness in dealing only with phenomena that have evidence in the scientific sense. But this is not satisfactory to me regarding my own soul.

Yes, I believe that my soul has existed for literally billions of years in the past in the universe, and this is my second life on Earth as a human being.

4. RETURN TO THE QUESTION OF STATISTICAL INFERENCE

Statistical inference about the true value of the variable v in the theory A from measurements v_1, \dots, v_N is not a trivial issue of taking the empirical mean.

Let us consider the problem of taking the mean. Let

$$\bar{v} = N^{-1}(v_1 + \dots + v_N)$$

This will produce an estimate of truth for cases of symmetric noise distributions, for example

$$v_j \sim v_{true} + \varepsilon_j$$

with

$$\varepsilon_j \sim N(0, \sigma^2).$$

In this artificial situation, we have:

$$|\bar{v} - v_{true}|^2 \sim \sigma^2 \chi^2(df = R(N))$$

Then we can do statistical inference here from $\chi^2(df = R(N))$ distribution. In particular we can take some specific value such as

$$q = \frac{1}{50}$$

and even though we have no access to v_{true} we can claim

$$P(|\bar{v} - v_{true}| \geq q) = P_{\chi^2(df=R(N))}(|Z| > q)$$

This leads to probability estimates for whether mean represents the true value. I won't go through tail estimates on χ^2 because the substantial issue is evaded by assuming $\varepsilon_j \sim N(0, \sigma^2)$.

There is no guarantee that the noise is Gaussian or symmetric in general and in fact it is in overwhelming number of cases not symmetric or Gaussian.

5. PROPOSED REMEDY

Our proposed remedy is to assume that errors in Social Science are always Generalised Hyperbolic Distributions and seek the appropriate replacement for χ_d^2 distributions. We need an ability to push forward the analogy of the normal theory and then we can use tails of new exact distributions for statistical inference for Social Science.

6. MONTE CARLO EXAMINATION

```
# Suppose the data x_1,...,x_N
# come from ghd(mu,theta).
# We do not know the actual mu at all
# Our task is to compute an estimator
# mu0. What is the confidence interval
# for mu0 in ( mu - a, mu + a)?

library(ghyp)
confint.ghyp<-function( x ){
  n<-length( x )
  fit.x <- fit.ghypuv(x,mu=2,sigma=sd(x))
  fx<-coef(fit.x)
  mu<-fx$mu
  print(summary(fit.x))
  M<-500
  tcount<-0
  error.dist<-ghyp(
    mu = 0,
    lambda=fx$lambda,
```

```

        alpha.bar=fx$alpha.bar,
        sigma = fx$sigma,
        gamma = fx$gamma)

tail<-qghyp( 0.95 )/(n/55)
f <- sum((x - mu)^2)/n^2
for ( r in 1:M ){
  gsmp<-rghyp(n, object=error.dist)
  e<-sum(gsmp^2)/n^2
  diff<-abs(f-e)
  if (diff> tail^2){
    tcount<-tcount+1
  }

  if (r%10==0){
    print(paste(r,diff,tcount/M))
  }
}
width<-tcount/M
list(estimate=mu,low=mu-width,high=mu+width)
}

```

6.1. Real Variable Test.

```
V<-as.numeric(unlist(na.omit(wvs7[, "Q76"])))
```

With this code, we obtain for a categorical variable fit the following.

```

> gg
$estimate
[1] 6.845667

$low
[1] 6.829667

$high
[1] 6.861667

```

The tail is $p = 0.016$ and the width is $\Delta x = 0.032$ for the confidence interval, and $\mu = 6.846$. We want to interpret this as the true value is within $\Delta x = 0.032$ of the estimate. This is an exciting development for a theory of statistical inference.

7. LEGALESE

These are for ordered categorical variables with numerical values. I don't want a large number of accountants demanding that Generalised Hyperbolic fits are done on categorical variables with values ("crayon", "pencil", "ink", "happy", "sad") and others. Our interest is in analysis of categorical variables with an ordering.