

Statistical Inference and Data Analysis: G0P75b

Lael John

October 13, 2022

Preface

This set of notes deals with SIDA, my second course in statistics and probability. I'm looking to solidify concepts in statistics, and gain some insight into thinking about the correct statistical model(s) to use in a given situation. These notes are also going to be as detailed as I can make them, because they also help in filling in the gaps in my knowledge of probability and statistics. If it seems like I'm stating the obvious, feel free to skip ahead.

Contents

1	Statistical Models and Estimators	7
1.1	Probability Theory	7
1.1.1	Random Variables and Vectors	8
1.2	Mathematical Statistics	9
1.2.1	Types of Statistical Models (based on parameter space)	10
1.2.2	Statistical Inference	10
1.3	Random Sample	11

Chapter 1

Statistical Models and Estimators

What are these things? Firstly, statistics deals with stochastic (seemingly random) data, to try and better understand what is happening/why such data was generated in the first place. There's two major tools at a statisticians disposal

- Probability Theory: Gives us a framework/way of thinking about random phenomena, allows us to study interactions between sets of random events
- Random Samples: Gives us a foundation upon which to apply our probability theory knowledge.

1.1 Probability Theory

Probability theory must be studied in some "space" so to speak. We deal with classes of events and how those "events" or classes can potentially distort each other. Further, because we deal with such an arbitrary definition of what events are, it becomes necessary to first transport every event to some portion of the REAL NUMBER interval $[0, 1]$, allowing us to then begin working from a numerical foundation.

That being said, we denote a probability space as (Ω, \mathcal{A}, P) where the first element denotes the space of all outcomes, the second element giving us what

we will later come to know as a σ -algebra, and the last giving us a probability function/MEASURE.

Remark. Note that when talking about the universe of all outcomes, no mention is made of the fact that those outcomes may or may not be possible. That falls into the realm of our second element, the σ -algebra. This term seems to deal with a class of subsets, each of which can be measured?

Thoughts. The concept of a "measure" is one that is slightly abstract. Such a function returns a sort of volume? Or assigns a number to a certain region of the space we're considering, but that number also seems to be determined by a function that's already on that space (like the definition of the Lebesgue measure.)

1.1.1 Random Variables and Vectors

Now when we talk about a function, $X : \Omega \rightarrow \mathbb{R}$, that can be considered a random variable if it is a "measureable" function. Again, take this at face value, but the underlying principle seems to be that elements of a σ -algebra on(in) our target space are pulled back to elements of our σ -algebra in the probability space.

NOW, because X is a random variable, mapping the probability space to the set of measurable subsets in \mathbb{R} , we can also see that P induces P_X onto the reals, where $P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega | X(\omega) \in B\})$. Here the second of our 3 equivalent expressions shows us the reality of what this induced probability really is. P gives us the measure of the set that maps to B in the target space. The map that gets us to the target space is X , so that map that we need to use to get the set in the domain is precisely X^{-1} . Because we have access to our random variable now, we can begin to talk about other functions that involve the random variable, namely

1. Cumulative distribution function: $F_X(x) = P(X \leq x) = P(\{\omega \in \Omega | X(\omega) \leq x\})$
2. Density function: $f_X(x) = \frac{dF_X(x)}{dx} = P(X = x) = P(\{\omega \in \Omega | X(\omega) = x\})$ (Note that this only works if X is absolutely continuous? Will need to review why this is the case.)
3. Moment Generating Function: A holdover from using mechanics terms within mathematics (curse you Newton), $M_X(t) = E[e^{tX}]$, i.e. the

weighted average of all the e^{tx} for all x , and a particular t . Note also that this function may in fact, not exist at all, but the next one, our *characteristic function* will, for all random variables.

4. Characteristic Function: Similar to the above except $\phi_X(t) = E[e^{iXt}] = E[\cos(Xt)] + iE[\sin(Xt)]$

A random vector therefore is basically a collection of random variables from the probability space to the real numbers. Here the probability measure that is introduced becomes the measure of the intersection of the preimages of all the values given by each random variable. Similarly the cumulative distribution function can be defined, though now when we talk about the density function, we take the derivative of the CDF with respect to all our random variables, not just one. (NOTE: some property of higher order derivatives needs to be fulfilled in order for this to happen, needs a little more research). The moment generating function and characteristic functions again, remain the same, with their respective modifications.

Conclusion

In essence when dealing with probability theory, we have a well defined measure that we can apply onto our universal space, and we study the property of that space, with its sigma algebra all while using our trusty known measure. THIS DOES NOT HAVE TO BE THE CASE WITH MATHEMATICAL STATISTICS.

1.2 Mathematical Statistics

When we begin to talk about mathematical statistics, we begin simply with our universal set of data, and we assume various probability models (or spaces in our previous terminology). Thus our statistical model becomes either one of the following

- $(\Omega, \mathcal{A}, \{P_\theta | \theta \in \Theta\})$
- $(\Omega, \mathcal{A}, \{F_\theta | \theta \in \Theta\})$

Here Θ denotes the parameter space, within which θ varies, and P, F correspond to various probability measures or density functions respectively.

1.2.1 Types of Statistical Models (based on parameter space)

One way of classifying our statistical models is with the relative size or properties of our parameter space.

1. Parametric statistics: When Θ has finite dimension (i.e. when it can be thought of as a vector space, each n -dimensional vector corresponding to n different parameters to be used to create the current probability measure in question).
2. Non-parametric Statistics: Here, stupidly almost Θ must have an infinite dimension. Therefore it makes no sense for us to even think about specifying a θ . In this case, the different measures correspond to different sequences (countably infinite) or functions (possibly uncountably infinite).
3. Semi-parametric: What you get when you decide that some parameters can be specified in finite dimensions, but you need to append an infinite dimensional parameter anyway.

1.2.2 Statistical Inference

Inference, or figuring out what θ seems to have generated the data that we're looking at. Looking at what we've covered already, we're studying a particular value of our statistical model in consideration, knowing that we already have some data to work with. In essence, given a large family of models, we're trying to find a particular parameter that seems closest to reality. (This is a little confusing, given that when we want to talk about reality, we are also using another probability function? even if it's an approximation??? You're basically approximating an approximation, what is happening in the world.) To end up being successful in our process of inference, we must necessarily make some estimates about the state of our data. Those estimates (or estimators that generate our estimates) can be classified simply into the following categories:

- Point Estimators: Where we're trying to find a relatively good approximation of θ .

- Confidence intervals: Trying to determine some subset of Θ , that with a reasonably high confidence, we can say θ lies.
- Hypothesis test: Basically partition the space into two, and classify θ as lying in one part of the space, or its complement.

1.3 Random Sample

A random sample, is a collection of *independent* random variables, all having the same distribution as X . These are then called individually independent random variables. (Essentially a tuple, each using the same map on the coordinates.) The way one writes a random sample $(X_1, X_2 \dots X_n)$

A *statistic* is now a function from a random sample $(X_1, X_2 \dots X_n) \rightarrow T(X_1, X_2 \dots X_n)$. Now if we go from an n-dimensional universe of outcomes (n different universes of outcomes) to a simple real number, then $T(X_1, X_2 \dots X_n)$ is simply a random variable. If however, we go to p-dimensional real space, then each function $T_i(X_1, X_2 \dots X_n)$ from $i = 1 \dots p$ is component of $T(X_1, X_2 \dots X_n)$, where each is a random variable, resulting in a random vector.