Statistics for Information Intensive Agriculture

Anil Gore

Cytel Statistical Software and Services Pvt. Ltd. Pune

July 2018

1. Abstract

This paper proposes that productivity growth in rain fed agriculture is possible with more intensive use of available information. This may be a way out of the productivity plateau experienced since the green revolution. Some illustrative examples are discussed to spell out the ideas. This is intended more to help students of statistics understand their subject as a tool of development, rather as a ready device to be put into immediate practice.

2. Introduction

My experience as a student of statistics was one of endless wait for someone to tell me how all the theory stuff was useful in my own country. My teachers had no real interaction with any fields such as agriculture or industry in which the tools were supposedly to be used. So all illustrations of problems discussed in class seemed either bookish or alien. It is only after I became a practicing statistician myself that many things learnt as theory assumed a more meaningful form. So I propose to put down on paper some of my experiences hoping that they may help a curious student of statistics in understanding his/her own subject better.

In this essay I want to take up simple applications in agriculture. Of course statistics text books do give examples. But they do not make an impression. This is because the student has no perspective of this domain of application. So let me begin with some remarks about agriculture in India.

India’s population at the time of independence in 1947 was about 400 million. It was a poor and undeveloped country in which farm productivity was so low that the food available was quite inadequate. There were famines. Society was critically dependent on agriculture which depended on monsoon rains. Food grains were imported every year. The American government’s Public Law 480 was the difference between normalcy and starvation. It was a grim situation. Within the next 30 years or so, the picture changed. Today in 2018, our population is nearly three times the figure at the time of independence. And yet, not only are we able to feed ourselves, we have become food exporters.

So how was this achieved? There were several factors. Dams were built and area under irrigation went up. High yielding varieties of rice and wheat were adopted. Use of chemical fertilizers and pesticides went up. Together they were called the green revolution.

Like in most such cases the progress had a flip side. One is inequality. Only a small fraction of farmers participated in this upsurge of productivity. Majority with no access to irrigation did not participate and did not benefit. Environmental impact of these changes was often adverse. There was water logging, soils became saline, rivers and lakes were polluted. There was loss of diversity. Many traditional varieties of crops simply disappeared. Some of them had features such as fragrance and taste (attractive to the consumer) and pest resistance, drought resistance etc (attractive to the farmer). In any case there was loss of gene pool. One last feature of concern to planners was that while population continued to grow (though now at a more moderate growth rate) productivity in green revolution seemed to reach a plateau. So, it seems that a sustained growth in total food production needs improvement in rain fed agriculture.

These farmers are not rich and cannot really pay for expensive inputs. Also the enthusiasm for chemical fertilizers and pesticides has waned now. So what is the way out? I will argue that the next generation input should be information. We have to see how information input can help farmers in making wiser decisions. So our illustrations will relate to some specific decisions that dry land farming involves and the innovation suggested will be based on better use of information. A key input needed is study or relationship between weather and crop yields. (for example see [1]).

3. Two conceptual examples:

These examples are conceptual in the sense that I will not present any data. They are not hypothetical. They are indeed about decisions that a farmer takes and about how the decisions can be made more effective using information and statistical analysis. First example is of a fruit crop and second of a grain crop. Both came up during discussions with experts in agriculture.

A horticulturist (cultivator of fruit trees) has an annual cycle of caring for his trees. One step in the annual cycle is trimming at the beginning of the season (say end of summer). Then with rains there is growth of new leaves, then flowering and finally formation of fruit set. A farmer can (within limits) pick his date of trimming. Our concept is to provide a suggestion for this based on analysis of local rain fall data. One risk in case of fruit trees is of getting rains at the wrong time. Typically, flowers are vulnerable to rain. If there are rains during the flowering season, yield goes down. So, can statistical analysis help dodge this risk? Generally the number of days from trimming to flowering and number of days to end of pollination are roughly known. So, if we pick a date for trimming, we know the date when flowering begins and when it ends. From history of local rainfall, we can calculate risk of rain in those days. This can be repeated for different trimming dates. A good decision would be a date with minimum risk of rains during flowering.

A farmer looks for a new high yielding variety of groundnut and decides to try it in the new season. In this case the sowing date is dictated by rain and soil moisture. As soon as there is enough monsoon rain and soil moisture level is adequate, the farmer sows the seed. The remaining time table is not in his hands. It is dictated by the biology of the variety. One important feature is the number of days to maturity. The new variety matures in 110 days while the traditional variety took 90 days. The longer wait is quite worth the while if the extra yield is substantial. So the farmer is hopeful to do better this time. Then disaster strikes. Just as the groundnut crop matures and farmer makes arrangement for harvesting, it rains. The field becomes muddy. Groundnut pods are underneath and have to be pulled out. If soil is wet, this extraction becomes difficult. A big chunk of the pods is left behind in the soil and essentially lost. Now this was not foreseen. The farmer ends up getting less not more yield. He did not face this problem with the traditional variety. One possible reaction is that farming is a risky business and gods did not favor him this time. If that is the final verdict, there is nothing much to be done. But in fact there is a possibility. What if the rainfall pattern is such that a 90 day variety is safe but a 110 day variety is vulnerable to harvesting problems due to rain? We can check this out. A locality may have a 100 year history of daily rainfall. For each year, we can decide which would be the sowing day (from the accumulated rain up to that day). Once a sowing date is decided, we can check the risk of rainfall at the harvest time. The latter will depend on the variety. So we will be able to estimate harvesting risk for each available variety. Some varieties may be unsuitable because of the local rainfall patter. Notice that rainfall pattern can vary from one locality to another. So, for each locality we will be able to screen and rule out some varieties. This needs to be done for each crop, each variety of each crop and for each locality. We are talking about a massive exercise of analysis regarding crop-rainfall relationship. This is what we mean by information intensive agriculture. According to us, such work is not done in India and provides ample scope for statisticians and for increase in productivity without chemical inputs.

Note that the proposed analysis is quite different from what is done today by the India Meteorology Department. Our idea is not to predict rainfall but use information on rainfall to improve farmer's decisions.

4. Pest Control:

Now let us take up some examples with data. In each case we should address a specific decision that a cultivator practicing rain fed farming takes and should try to improve the decision. Our first case study involves pest control in groundnut. The locality studied is Chitradurga district in the state of Karnataka. The issue is how best to control the pest 'leaf miner'. Let us first summarize the parameters. As stated earlier, sowing date is dictated by moisture accumulation in soil. Once that happens farmers sow seeds. After that pegs (as pods of groundnut are called) formation begins at the age 35 days and growth of these pods continues for 40 days. In this period if there is a dry spell of 15 days, the insect leaf miner attacks the crop. This dry spell is needed for promoting its population. Now we can go to the rainfall history data. We found that out of 84 years for which there was serviceable data, conditions suitable for leaf miner attack occurred in 58 years. So roughly there is 70% chance of leaf miner attack in any year. Once the population starts growing, it explodes and can wipe out the crop in 15 days. So what is the decision problem to be solved? Here are the options before the farmer: (a) Do nothing- this is the option for the poor farmer. He cannot afford the cost of pesticide. So either the crop is lost or there is some rain and the pest is wiped out. (b) Spray pesticide immediately- this is the option for the rich farmer. This may clear the pestilence. If by chance there is rain before the pesticide acts, then it may get washed out and spraying may be futile. (c) Wait for a few days. If it rains, that may eliminate the pest. Otherwise spray. This is a via media.

The various alternatives have to be compared in terms of costs and benefits. Costs are cost of pesticide (if used) and loss of crop. Benefit is the final harvest size. In calculating these, we need to input rainfall information. In particular calculating the probability of corrective rains is crucial. So again a thought experiment becomes necessary. For each of the 84 years with adequate data, we have identified 58 years in which pest attack seems plausible (at the end of a 15 day dry period). Now check when the next rainfall is observed.

Following model building exercise was done on this data set: random variable of interest is X: number of days between leaf miner attack and corrective rains. Assume that its probability distribution is geometric.

$$P(X=j) = p*q^{(j-1)} \quad j=1,2,...$$

This model fits well with estimated probability p = 0.14.

At this stage a new wrinkle becomes apparent. Does the chance of corrective rains change with time of attack i.e. days since beginning of Peg formation phase? Here is a table suggesting that it does.

Time of Attack (days)	Time of occurrence of correcting shower			Total
	Early (Within a week)	Late (After a week)	Too late (After 2 weeks)	
≤ 16	5	12	1	18
≥ 17	23	6	11	40
Total	28	18	12	58

It seems Chance of nature cure of attack (by rains) depends upon time of attack. Early attack: 5/18=28%.

Late attack: 23/40 = 58%. In view of this we propose an intuitive strategy namely "wait for a correcting shower if attack is late and use pesticide if attack is early". One may think of a strategy that waits for x days regardless of when the attack came and sprays only if there are no corrective rains in those x days.

5. Cost-benefit analysis of 4 strategies:

We need a loss function. It should lead to complete loss in 15 days. So we choose

Loss function L(j) = e.33*j where j is the number of days pest gets free hand, L(j) =% crop lost up to j days . Notice that L (j) ≈ 100 % if j ≥ 15. It turns out that waiting for 6 days is the most economic strategy. The table below refers to two sets of calculations. One involves use of geometric distribution model for rain while the second uses just the empirical distribution. Either way, waiting for 6 days provides a better outcome compared to any other strategy. This waiting helps reduce the ecological impact and also provides pesticide suppliers time to arrange a supply to the affected area. Again the method described here is rather specific to the particular problem. This is how it should be. The same method was applied to another locality (Anantapur district in Andhra Pradesh) and the answer came out to be similar. This suggests that there is some robustness in the solution. For details see [3].

Strategy	% Net Expected Income Using	
	Geometric model for corrective rains	Average from yearly data
No spray	77.22	73.02
Immediate spray	81.25	81.25
Decide on time of attack	78.47	78.47
Spray after 6 days	87.69	90.21

6. Fungus attack:

A farmer seems under siege through the entire life span of the crop. If there is a dry spell, the leaf miner attacks. If there is a wet spell, then a fungus attacks the crop. In this example we shall consider a particular fungus *Puccinia arachidis.* This fungus is supposed to have invaded India through the traditional route of invaders namely the Khyber Pass and slowly progressed from north to south. Potential loss due to this fungus is quite heavy. Farmers are known to have switched from groundnut to sunflower to avoid losses due to this fungus. So the decision problem here is how to protect the crop from this threat. Here we have access to some experimental data. In the planned experiment there were 60 experimental units. Crop was grown under varying weather conditions in pots. Fungus was inoculated at plant age 40 days. Response recorded was fungus severity every 10 days till plant age 120 days. There were daily records of weather including daily max, min temp, humidity, rainfall, sunshine hours etc. Analysis was done in two stages. In the first stage logistic model was fitted to fungal growth for each unit. Parameters r-growth rate and K- highest severity were estimated. In the second stage these two estimates were the inputs and the attempt was to study relationship between parameters and weather. The first step is rather straight forward. Second step also seems easy. After all we are trying to fit a regression. How difficult can that be? But a problem here is of having too many weather variables and only 60 data points. How many ? Hundreds. (120days X 5 weather parameters every day). So the problem is how to choose the 'best' subset? One way of reducing the number of predictors is to take response at say 10 days as input. It means weather data for first 10 days is now incorporated in this one number. I think a decision postponed for 10 days may not be a serious drawback. (You have to continually interact with potential users of your work to make sure that your assumptions are not ridiculous). In case of parameter K (which is the maximum severity of fungal attack) the result is nice.

K= f(severity on 10^{th} day, sun-shine hours on 8 to 12 days, Max temp on 10^{th} day, Min temp on 7^{th} day).

We get a good value of R^2 = 80%. On the other hand regression with growth rate as response is not successful. This may be because it is more dependent on the genetic constitution of the plant than weather.

What use can the farmer make of the regression equation for K?

When fungus is noticed in the field, one can use temperature and sunshine hours data for 3 days before and 3 days after to predict max severity. In other words the upcoming challenge to the health of the crop i.e. max severity can be anticipated 6-7 weeks ahead of time. (for details see [2]).

How is this useful? Agriculture experts see two uses. One is to prepare for the attack with prophylactic spray by timely scheduling this preventive measure. Second use is to ensure that the plant is healthy and strong to fight with the attacker, by making up micro/ macro nutrient deficiencies if any.

7. Conclusion

This note attempts to describe possibilities of enhancing productivity of rain fed farming using weather information inputs. These are ideas. Such methods are likely to be specific to a crop, to a variety of a crop and to a locality. Implementation of such ideas will involve very large interdisciplinary effort on the part statisticians, agronomists, entomologists, plant pathologists, agro-climatologists etc. This is a challenging vision but with great potential.

8. References

1. Predicting Jowar Yield Using Weather: Possibilities and Limitations(2003), In Proceedings of National Symposium on Statistical Methods and Its Application, Dept. of Statistics, Annamalai Univ.(Ed. R. Elangovan), pp232-243. With P. Arumugam and S.A.Paranjpe
2. Modeling progress of fungal attack on groundnut J. Ind. Soc. Agr. Stat. P.V. Sukhatme memorial volume LI, August-December, 1998, pp293-302. (With C.D. Mayee, Anita Shah S.A. Paranjpe).
3. Wait and see strategy for control of leaf miner in rain fed groundnut (1998) (with S.A. Paranjpe) Current Science, Vol. 74 No.4, p. 296-299.