LS 812 - Mathematics in Science and Civilization

Nov 3 2007:

Exploring Randomness: Delusions and Opportunities

Sources and resources:

Nassim, Nicholas Taleb *Fooled by Randomness,* Second Edition, Random House, New York.
Weldon,K.L. Everyday Benefits of Understanding Variability. Presented at Applied Statistics Conference, Ribno, Slovenia. September, 2007.
www.stat.sfu.ca/~weldon

Introduction: The discipline of statistics has origins in gambling, astronomy, and experimental science. Mathematics helped to clarify the properties of statistical methods and produced many useful developments: correct odds were worked out for gamblers, measurement errors were made allowance for in astronomy, and Fisher put experimental design on a firm basis. Underlying all these origins, the concept of "randomness" enabled researchers to describe "unexplained variation" unambiguously. Of course, this description was not real, only realistic. We had an abstract way of describing that part of measurement variability that defied attribution to specific causes.

Mathematics provided the tools to cope with "random" variation. Parametric models (formulas for relative frequencies calibrated by a few constants called "parameters") became the primary tool for expanding this idea of randomness to complicated applications. However, the advent of widespread computing power that has evolved over the last 50 years has changed how people study randomness: computer software in combination with these parametric models has provided methods of analysis that are accessible to a wide variety of professionals, not only mathematicians. In fact, recent trends in the development of graphical methods have moved some statisticians away from the parametric models, although the concept of randomness is still central to the proper analysis of data.

The discipline focused on coping with unexplained variation and randomness models is called "Statistics". In this talk, I will provide ten examples of how the discipline of statistics can help to reveal some surprising randomness phenomena, and for each of these examples, I will point out the general concept illustrated by the example. Many of these examples use a technique called "simulation" or sometimes "Monte Carlo simulation", and this technique requires the parametric models invented by mathematicians more than half a century ago. We will show that the mathematics is essential for research in statistics, but that it is possible to appreciate the implications of statistical thinking without mathematical detail: the ten examples provide information that is useful in everyday life.

1 – When is success "good luck"?

The book "Fooled by Randomness" explores the wide variety of situations where "success" seems to be attributable to superior intelligence or vision, but where closer analysis reveals "good luck" to be the true cause. The sports scenario used here is just an easy one to comprehend: the inclination is to attribute a winning streak to team quality, but the analysis shows that the impact of luck is under-rated.

Teams in a league are updated yearly to try to balance the quality of the teams. The reason for this is economic: the excitement that fans expect will depend on the game being fairly close in score. The biggest turnouts are for games for which the outcome is the most in doubt. In this situation, a reasonable model for the league play is to assume that each game is a 50-50 game. The surprising thing that a simulation shows is that this assumption still provides for a wide range of league points from the top to the bottom of the simulated league. Comparing the spread of league points in the 50-50 simulation of a season, with an actual league outcome, confirms that much of the apparent difference in the success of the various teams is attributable to "luck".

2 - Order from Apparent Chaos

In the previous section we showed that apparent order can actually be an illusion of chaos. In this section we show that apparent chaos can actually reveal a surprising orderliness. I collected my gasoline consumption for every fill-up over a five year period. (This odd behavior may be explained by my odd profession: statistics!) The graph shows no apparent pattern. A statistics student might try to fit a line or a quadratic curve or perhaps a higher order polynomial to check for patterns in this kind of data. This approach just produces a horizontal line, and a conclusion of "no interesting pattern". However, with modern software which makes use of graphical methods to analyze data, it is not hard to find a pattern. The question then is, what causes this pattern? It is surely interesting in the context of the data to know what factors are associated with higher or lower gasoline consumption. A technical point of the example is that graphical methods are very powerful, and we should not rely on fitting simple parametric curves when the motivation for them is absent. A larger point is that statistical methods exist that are both simple to explain and have practical uses.

We also use this example to talk a bit more about "smoothing". There is always a trade-off between the amount a smoothing and the amount of detail one wants to extract. We argue that the choice is subjective, and depends on the context of the analysis, and that subjectivity in statistical analysis is a fact-of-life. Statistics can lie but so can words! Always check your sources for bias.

3 – Utility of Averages

In the smoothing of the previous example, we hinted at the usefulness of averaging. In this unit we provide an example of employing averaging to reveal an important and somewhat surprising fact about "risky" investments. The well known principle of diversification - "Don't put all your eggs in one basket" - can be refined. However, we need to get out of the farmyard for a more quantitative example. The risky company described in the talk had a poor chance of making money but on average it did make money. A portfolio of several such risky companies whose fortunes were not related (not in one industry for example) can be shown to make money very reliably.

(Technical Note: Averages of independent outcomes have less variability than each outcome in the average. The factor relating these two variabilities is the square root of the number of outcomes.)

The investment result is surprising because "conservative" investors shy away from companies that have a good chance of going bankrupt. The result is useful because it suggests how to make money from an underutilized source.

4 – Industrial Quality Control

Imagine you are responsible for a shoe manufacturing company. There are several materials required, several designs, several sizes, several colours and finishes, ...in short, a very complex manufacturing process. In the old days (pre-1950) quality control was done on the finished products. A more modern approach is to measure everything at every step of the process. Of course this produces a huge amount of data, and managers cannot be looking over everyone's shoulder at all times. There is a principle called "management by exception" where only the exceptional cases are examined in more detail. A simple implementation of this is the control chart – this was promoted in Japan in the 50s with spectacular results – Japan's reputation for shoddy goods turned into a reputation for high quality goods over a fifteen year period.

The control chart is a way of signaling when a measurement is unusual – that is, outside of the normal range of values. The timing of unusual values often helps to pinpoint the cause, and the remedy usually results in less variability in future. This continuous improvement by reducing variability is a key to the success of Japanese industry, and is now being widely adopted worldwide.

The surprising thing here is that a simple device like a control chart could have such momentous consequences. The statistical principle underlying the control chart is that unusually large deviations often cast doubt on the stability of the underlying process. (Technically, this is the logic of "hypothesis testing").

5 – A Simple Law of Life

When classes of items are compared for size, there is often a predictable relationship in the relative sizes of the items. Zipf's Law is not really a "law" in the sense that it always occurs, but it is frequently observed empirically. The origin of Zipf's law is in the relative frequency of words in written English, but it has also found application in comparison of urban populations, company sizes, and internet activity. The surprising thing is that an apparently chaotic social process would bring about such regular outcomes. The usefulness of this is only that it provides a way of describing a size relationship. Of course it does suggest that there must be some rational explanation for the regular outcome, and there is lots of research into this question. Try searching for Zipf's Law on the internet.

6 - Obtaining Confidential Information

If you want to know the proportion of a class are regular marijuana users, they will likely not answer, or at least not answer truthfully, unless they are assured that their response will not be tied to their name. One way to protect respondent ID in responding to a sensitive question, is to use the randomized response technique:
1. Ask the respondent to toss a coin and to privately observe the outcome
2. If the outcome of the coin toss is a head, then the respondent should answer the sensitive question with "Yes" or "No" as appropriate. But if the outcome of the coin toss is a tail, the respondent simply answers "Yes" to the sensitive question.

When the responses are recorded, even if the name is attached, there is no way for the surveyor to know if the particular respondent is answering the sensitive question, or has simply got a tail on the coin and is not answering the sensitive question.
So suppose we have a class of 100 students, and 60 of them answer "Yes". We estimate that about 50 of them will have tossed a "Tail" and are answering Yes for this reason. But then about 10 are answering Yes to the sensitive question, out of the 50 or so who tossed "Head". Ergo, the proportion who say they are regular users is 10/50 or 20%.

Note that, of those that answer Yes, only 10 out of 60 are answering Yes to the sensitive question, so the true response of an individual answering Yes is well camouflaged.

I did this in-class survey with my 2002 STAT 100 class, which did contain about 100 students. The Yes answers were done with a show of hands! There may have been some who did not answer either question. The effect of this would be to bias my estimate downward.

The surprising thing here is that confidential information could be obtained even though responses were tied publicly to respondents. The more general lesson is that probability can have practical uses such as preserving confidentiality while still recording useful statistical information.

7 - Survival Assessment

"Survival Analysis" is a sophisticated technique usually used to compare the outcome of various treatments of life-threatening diseases. However, the outcome "death" need not be the one of interest, and the context need not be medical. As an example, consider the prospects of student drivers: what is the chance that a student will be involved in an automobile accident in the next month?

I again used my STAT 100 class (100 students) for this exercise. I asked the students to write on a slip of paper the answer to two questions:
1. In what month and year did you receive your first driver's license?

2. Have you been involved in an automobile accident (as driver or passenger)?

Is this enough information to estimate the chance that a student will be involved in a first accident in the next month?

The students did not identify themselves, and it was expected the information would be reliable. The objective initially is to use the data to estimate the relationship between the time-of-risk and the chance of having had an accident. We divided the data into time-of-risk intervals of 6 months, and recorded the proportion who reported having already been involved in an accident. The result was as shown in the talk, a straight line with slope of about 0.01. In other words, each additional month of exposure produced an increased chance of 1 percent that the individual would have been involved in an accident. For an individual that has been accident-free so far, we might estimate the chance of a first accident in the next month to be 1%.

Is this an example of "survival analysis"? Yes – but here "survival" means being accident-free. We are not comparing medical treatments for life-threatening diseases, but we could easily do a similar analysis for a different group of individuals and make that comparison.

There are many assumptions in this analysis, and we will not review them here. The point is that even minimal data acquisition can produce useful information when properly analyzed. (For more discussion of this example, see the notes at www.stat.sfu.ca/~weldon for STAT 100 on Nov 11, 2002)

8 – Lotteries: Expectation and Hope

This example just shows two things:

1. Lotteries sell hope rather than realistic chances to win.
2. As an investment, public lotteries are poor vehicles, even for the lifetime participant.

Lotto 649 for example, has a jackpot awarded to one set of numbers out of approx 14,000,000. If a person bought 10 tickets every week for 60 years, their chance of winning this jackpot would be about 1/5 or 1 %, and they would have spent about $31,000 for the opportunity. If they are clever, they will chose numbers that others would not choose so that they would not have to split the prize, but actually the chance of this being a problem is pretty tiny!

When a person buys a lottery ticket, it seems not totally illogical to think about what the person would do if they won the jackpot, and this is a pleasant activity for some people. So it may be that a lottery ticket has value other than its investment value.

The surprising thing about public lotteries is how small the average return is, and also how rare the large wins are. The general lesson here is that a simple understanding of probability allows a rationale assessment of the situation.

9 - Peer Review: Is it fair?

At SFU I served on a committee to assess potential student applicants that did not have the usual grades to enter SFU – it was called the Diverse Qualifications Admissions Committee. Each applicant was reviewed by two members of the committee who would bring a recommendation to a meeting of the full committee. In the early days of this committee, the decision for an applicant was largely a result of the particular reviewer who happened to be assigned to them. This did not seem fair and procedures were revised to eliminate this unfairness. However, the same unfairness persists in the treatment of submissions to academic journals: two referees are assigned to recommend publication or not. The simulation reported in the talk shows that referee variability – in other words the varying tendency of a referee to approve papers – adds a lot of unfairness to the review system. While accepted papers tend to be of high quality, many papers of equally high quality are rejected.

10 - Investment: Back-the-winner fallacy

When one looks at the annualized performance of mutual fund managers over several years, the rates of return typically fall in the 5 to 15 % per annum range. We can simulate this outcome by using a daily random walk model with the probability, p, of a day-to-day increase being in the range .54 to .56. "p" is a true measure of the quality of the manager.

We simulate the experience of 100 managers by assigning them a random p in this range – then we pick the "best managers" as indicated by their 5 year performance and, using the same p –value for each manager, simulate another 5 year experience. We then compare the performance of these "best managers" over the second five years with the "other managers". It turns out that the difference in performance of the "best" and "other" managers is very small. The reason is that the variability in performance over a 5 year period swamps the quality difference as measured by p. In other words, the managers chosen by performance are not really much better than the other managers. So chosing managers based on past performance is not a very useful strategy.

Conclusion

I have attempted to show how a little statistical knowledge can reveal information that is not widely known and yet is useful for real life. I have also tried to provide some appreciation of the methods used in deriving this information. If anyone would like clarification at some later date, I am easily contacted by email at weldon@sfu.ca.