Samples from the history of probability and statistics

1 Censuses

First came data collection in the form of a census. Ancient societies would do a census of land, or of people, or of property, or of certain people (for example, free citizens only), or of certain kinds of land (for example, farmland), or of certain kinds of property; or of various combinations. The first census of which we have a record was in Egypt around 3050 BC. Several censuses are mentioned in the Bible, in both the Old and New Testaments. Here's an extract from the English *Domesday Book*, a census written up in 1086:

> "In (North) Allerton there are 44 carucates[1] of land taxable, which 30 ploughs can plough. Earl Edwin held this as one manor before 1066, and he had 66 villagers with 35 ploughs. To this manor are attached 11 outliers [i.e., other estates, listed in the original; we won't list them here]... Now it is in the King's hands. Waste. Value then 80 pounds. There is there, meadow, 40 acres; wood and open land, 5 leagues long and as wide."

The first thing you learn these days about data analysis is to have a clear purpose (or clear purposes) that guides your gathering of data. The second thing is to describe your data clearly.

By modern standards the *Domesday Book* is quite jumbled and unclear. Do the "66 villagers" include women? children? old people? Why are estates counted but not the nobility? (Surely some estates had more than one noble in residence). Why count ploughs but no other kind of property? Presumably "taxable land" means farmland. What is worth 80 pounds? Does "waste" refer to all the outlying estates? Or the meadow, wood, and open land? How do you make sense out of all this raw data?

2 Data analysis

In 1532, the English government began gathering together parish records on deaths in order to keep track of the plague. The data was not accurate — a parish clerk might miss a week, and then make up for it the next week by reporting two weeks' worth of data as one; families might not correctly report the cause of death out of fear of being shunned; and, because medicine was far from an exact science, an honest report could simply be wrong.

In the mid 17th century, John Graunt systematically studied many decades of this data, looking for patterns and trends, and using what we would now recognize as statistical processes to make conclusions about the data. For example, he noticed that in 1625 the number of reported deaths not due to plague formed a sharp spike in the data. He found that unbelievable, and concluded that there were more plague deaths in 1625 than reported.

3 Economics

As the European economy got larger, more complex, and more capitalistic from the Middle Ages on, it became necessary to estimate probabilities. For example:

[1] Wikipedia: "an area equal to that which can be ploughed by one eight-oxen team in a single year (also called a plough or carve). Approximately 120 acres." Not sure how many square khet this equals.

- You're a marine insurer: shipowners pay you a premium up front and you agree to pay them the value of the ship if it is wrecked or captured by the enemy. How much do you charge? (It depends on how likely the ship is to meet disaster.)
- You're a government treasurer raising money by selling annuities (an investor gives you cash, and you pay the investor a fixed amount per year for the rest of his life). What should that amount be? (It depends on how long the investor is likely to live.)

Without an accurate theory of probability underlying these kinds of decisions, it's easy to lose a lot of money (and many insurers and treasurers did).

4 Astronomy

While much of modern statistics has come about through such fields as quality control in manufacturing (e.g., the t-distribution was discovered by the statistician William S. Gosett who worked for Guinness Brewery), agriculture (the USDA was established in 1862, and its Division of Statistics was formed one year later), public health (the work of Florence Nightingale), or social science (especially psychology), the first real statistical problem that was widely studied was the problem of astronomical measurement. Even in the second century BC, astronomers knew that their measurements were not precise. They knew that different measurements of the same event would be different because of conditions they could not control, such as the vagaries of the instrument, or of the atmosphere, or of the weather. Different astronomers dealt with this situation differently: some used what we now call the mean, others the median, others grouped data, or resorted to using ad hoc formulas (which they often didn't report)... It wasn't until the mid 18th century that the mean was the standard method of summing up repeated observations, and even then there was controversy over whether you shouldn't just take one observation and stick to it — why complicate your thinking with all these other observations? what could they really add to our understanding?

5 Gambling

Cardano (the same guy involved in the controversy over the cubic) wrote a book called Liber de Ludo Aleae (The Book of Games of Chance) around 1550, though it was not published until 1663. Cardano, Tartaglia and others argued about this problem (quoted in Burton, p.400):

> A team[2] plays ball so that a total of 60 points is required to win the game and the stakes are 22 ducats. By some incident, they cannot finish the game and one side has 50 points, the other 30. What share of the prize money belongs to each side?

In other words, how do you determine the probability of either team winning? The problem is not entirely well-defined, but it is tougher than it looks, and all the solutions proposed were incorrect. Around the time that Cardano's work was published, Fermat and Pascal were dealing with the queries of the aristocratic (and somewhat intellectual) gambler Chevalier de Méré. For example: "Two players alternately flip a fair coin. Whoever is the first to get heads wins. What are the odds that the first player wins?" Or, "I offer you an even-money bet that if I roll four dice, I will roll at least one 6. Should you take the bet?" (Today we'd consider these problems relatively elementary, but probability still can be confusing, particularly when there are issues of information involved. An infamous recent example is the Monty Hall problem.)

This was the birth of probability theory: essentially all the basic rules of probability theory came from this work, as well as quite a bit of advanced probability theory. Probability theorists find themselves calculating

[2] This probably means "Two teams".

the binomial distribution, used to figure out what you could expect in a repetition of n trials when there were only two possible outcomes (for example, flipping a coin n times). There is an exact formula for the probability of getting exactly h heads, using a binomial coefficient: $\frac{1}{2^n}\binom{n}{h}$.

If n is small it's easy to calculate. But if n is large this is hard. So people tried to find good ways of approximating the binomial distribution. De Moivre, in 1733, proved that as n and d get large, the probability of getting exactly $\frac{n}{2}+d$ heads in n flips approaches $\frac{2}{\sqrt{2\pi n}}e^{-2d^2/n}$. (This looks like a complicated formula, but it is really much easier to calculate than the exact value, $\binom{n}{n/2+d}/2^n$ for, say, $n = 2008$.) Therefore, the probability of getting between $\frac{n}{2}$ and $\frac{n}{2}+d$ heads in n flips of a coin is

$$\frac{4}{\sqrt{2\pi}}\int_0^{d/\sqrt{n}} e^{-2y^2}\,dy.$$

If you've had a calculus-based statistics course, you recognize $\frac{2}{\sqrt{2\pi n}}e^{-2d^2/n}$ as one of the family of curves we now call the normal distribution, and you may remember that part of the course involved using the normal distribution to approximate the binomial distribution. Given the centrality of the normal distribution to the rest of statistics this might have struck you as somewhat quaint, but in fact it is exactly the effort to approximate the binomial distribution that is the origin of the normal distribution.

6 Handling errors in scientific observations

Going back to astronomy, mathematicians were trying to figure out how to describe the errors of observation they saw in astronomical data. Some principles were clear, for example:

1. Small errors are more likely than large errors.
2. For any real number ϵ, the likelihood of errors of magnitudes ϵ and $-\epsilon$ are equal.

The goal was to find a function $y = \phi(x)$ so that the probability of an error between r and s was $\int_r^s y\,dx$. This automatically adds a third principle: the total area between the curve and the x-axis must be 1.

Using these principles, Laplace proposed two curves which had the disadvantage of either not being differentiable at 0 ($y = \frac{m}{2}e^{-m|x|}$) for some constant m, or having a vertical asymptote at 0 ($y = \frac{1}{2a}\ln\frac{a}{x}$ where $-a \le x \le a$). This meant that these principles were not sufficient to determine the error curve.

Gauss added a fourth principle: Given several measurements of the same quantity, the most likely value of the quantity is their average.

Using these four principles he determined that $\phi(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$, where σ is the standard deviation. (Actually, he didn't quite do this, since he didn't have the notion of standard deviation. Instead, he had a quantity h which he thought of as the "precision of the measurement process"). This is, of course, yet another normal distribution.

Gauss did not do this in an abstract context. On January 1, 1601, an astronomical object (in fact, the first asteroid to be noticed by humans) was discovered by the Italian astronomer Giuseppe Piazzi, who named it Ceres. To ascertain its orbit, many people observed it and recorded their observations. But six weeks later Ceres disappeared behind the sun. Where would it reappear? Gauss suggested searching an area of the sky that differed from the one most astronomers predicted, and he was right. It was this error curve that enabled him to make the prediction.

When Gauss published his work in 1809, he claimed that his fourth principle depended on the method of least squares — what we call least squares regression. This was first published by Legendre in 1805, although

Gauss claimed that he had known about it since 1795.[3] This method is used to find a straight line that best fits the data, and since it's very technical, that's all I'll say about it.

The error curve has the property that its maximum height is at $x = 0$. Generalizing the formula to account for maximum heights elsewhere gives the family of normal distributions: $N(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ where μ is the mean and σ the standard deviation.

7 Buffon's needle problem

Take a piece of lined paper in which consecutive lines are at a fixed distance d. Throw a needle with length $\ell < d$ onto the paper. What is the probability that the needle crosses one of the lines?

Answer: Suppose the needle (if extended) makes an angle of θ with one of the lines, where $0 \leq \theta \leq \pi$. Then the probability that the needle hits a line is $\ell \cos\theta/d$. Averaging this over all θ we get

$$\frac{1}{\pi} \int_0^\pi \frac{\ell \cos\theta}{d}\, d\theta \quad = \quad \frac{\ell}{d\pi} \int_0^\pi \cos\theta\, d\theta \quad = \quad \frac{2\ell}{\pi d}.$$

Pierre-Simon Laplace realized that you could use this to approximate the value of π: throw a needle many, many times (say N) and record how many times (say C) it crosses a line. It should be the case that $2\ell/\pi d \approx C/N$, so that

$$\pi \approx \frac{2\ell N}{Cd}.$$

In 1853, Johann Wolf tried it with a needle of length $\ell = 36$ mm and lines that were $d = 45$ mm apart. He tossed the needle $N = 5000$ times, and recorded $C = 2532$ hits. This must have been incredibly boring, but he ended up with the approximation

$$\pi \approx \frac{2 \cdot 36 \cdot 5000}{2532 \cdot 45} = 3.15955766\ldots$$

which is not great, but not too bad. (Was Wolf unlucky? How many more times would Wolf need to toss the needle to get an answer likely to be accurate to two or four or six decimal places? What does "likely" mean here?)

In 1901, Mario Lazzarini claimed to have redone the experiment with needle length $\ell = 25$ mm and line distance $d = 30$ mm; out of $N = 3408$ trials, he reported 1808 hits. This would give the fantastic approximation

$$\pi \approx \frac{2 \cdot 25 \cdot 3408}{1808 \cdot 30} = \frac{355}{113} = 3.14159292\ldots$$

However, there is strong evidence that Lazzarini's report is a hoax. First of all, the approximation $\pi \approx 355/113$ had been around for a long time; Zu Chongzhi had discovered it around 500 CE. Second, it's suspicious that Lazzarini picked a weird number like 3408 for N. Third, N.T. Gridgeman (1960) and L. Weber (1994) showed (using probability and statistics — what else?) that it is nearly impossible that Lazzarini got that lucky. (Gridgeman calculated that in order to be 95% sure you have approximated the first digit correctly, you need $N = 10000$ throws.)

The students of Math 410 conducted this experiment on April 24, 2013 by throwing toothpicks at the tiled linoleum floor of 301 Snow. The toothpicks were measured as $\ell = 2.61$ inches in length and the tiles are $d = 12.0$ inches on a side. Of $N = 450$ toothpicks thrown, $C = 78$ hit a line, which produces the estimate

$$\pi \approx \frac{2\ell N}{Cd} = \frac{2 \cdot 2.61 \cdot 450}{78 \cdot 12} \approx 2.51.$$

Perhaps we should run the experiment a few more times!

[3] This tended to happen a lot in the 19th century; X would prove a theorem, and Y would discover that Gauss had had the idea earlier.