A short piece on randomness

Meteorology is a probabilistic science, meaning that weather predictions will not always match up to the actual weather that we get. There are good reasons for this. A good forecast will give some guidance as to how probable its predictions are. An example of this is the rainfall prediction of the Met. Office web forecasts. Probability is a way of talking clearly about events that have a degree of randomness, so let me say a bit about randomness. This piece is not intended as an introduction to probability theory. It's just a note relevant to probabilities in forecasting.

Randomness of independent events

I'll start with a digression. Some randomness is caused by the variability in the way we make measurements. Tossing a coin high in the air and saying which side it lands on is the archetypical example. The amount of spin we give it and the height it goes to varies enough from throw to throw that the result is allegedly unpredictable. There is an equal chance of heads or tails in every throw. An important feature of this and many other random systems is that each result is independent of any earlier results. If there have been 5 heads in a row, that doesn't make it any more likely that the next one will be a tail. You can't say a tail is 'due'. The same goes for the ball in roulette. If it has landed on a red number 5 times in a row, don't put all your money on the next landing being black. The probability it will be black is just the same for each throw.

Some events in nature are intrinsically quite random. Let's take a physical example. Suppose a weak radioactive source emits a gamma ray on average every 10 seconds. If 9 seconds pass without any emission it does not mean that another particle is due in the next second or so. The chance of getting one in the next second is just the same as if the 9 seconds without one hadn't passed. I read an article today in a reputable publication saying that the average time between big disturbances in the upper atmosphere caused by mass ejections from the Sun is 150 years. The author said that the last such event was 160 years ago, so one is now 'overdue'. This is rubbish. Mass ejections are very common. Their size is randomly distributed. The chance of getting this year the kind of big one discussed is just the same as it is every year, regardless of the 150 years without one.

Some events have a random component but are not completely random. For example if it takes about 2000 years to fill the magma chamber below a volcano then we may not be able to predict the eruption time to within a century but once the volcano has more or less emptied the chamber in an eruption we know it's unlikely to erupt again for a long time. If the San Andreas fault is known to slip on average every century then since the last big slip was in 1906 one can really say that a slip is 'due', though it may be this year or in 10 or 20 years' time. That's the random element. The randomness here is not whether the stress in the fault will build up but when it exceeds the slip threshold.

Randomness in meteorology

The examples above are intended to show that to have some idea of the role of randomness, one needs an idea of its cause. So, what causes the randomness that underlies the need to put probabilities into meteorological forecasts? There are at least three causes. First, the atmosphere and the interacting ocean system has the inbuilt randomness of a chaotic system. Small changes at one time can result in big changes later on. At ground level the atmosphere is continually buffeted by small changes that are caused by the rest of the world – waves breaking unpredictably at sea, trees bending unpredictably in the forest, mankind and animals going about their business, and so on. Even above ground, clouds form and evolve with random detail, perhaps blocking sunlight at random or changing the local distribution of water. The twinkling of stars at night makes visible random fluctuations in air density in the atmosphere. Add to all this the fact that phenomena associated with precipitation depend on nucleation, accretion and collision and this increases natural randomness of the weather. For example, simply having the atmosphere saturated with water vapour doesn't guarantee the formation of water droplets; cooling the atmosphere below 0°C doesn't guarantee that drops will freeze; having enough water in the clouds to make rain doesn't guarantee it will actually rain. The weather system has inbuilt randomness. The bottom line here is that meteorology will never have the predictability of mechanics. Predictability is an issue of scale, in space and in time. You won't know at Christmas whether your summer holiday weather will be good or not, simply because the weather system itself is indeterminate on that time scale.

A second source of forecast randomness comes from the forecast process itself. Forecasts are computer calculations based on the physics of an approximation to the real world. Approximations simplify. Sometimes the simplifications will introduce an error one way, sometimes another. For example they may result in predictions of too much rain sometimes and too little at other times. So the approximations in both assumptions and how the equations are solved effectively introduce an inaccuracy that appears as a random error in the forecasts. Meteorologists are very aware of this and part of the real improvements in forecasts in recent decades has been due to a reduction in this forecast randomness by the use of larger, faster computers that handle more realistic models and better numerical solutions of the equations of the underlying physics.

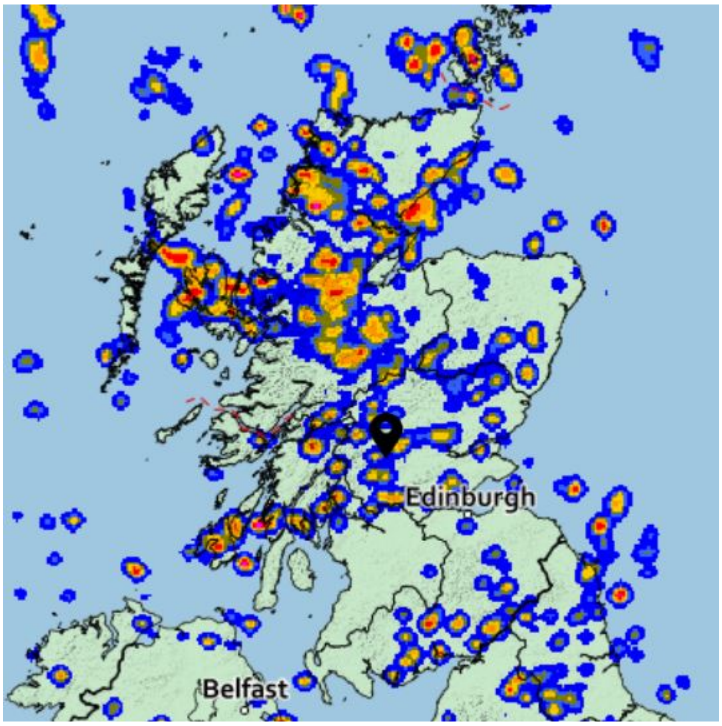
A third source of randomness arises from the fact that the computers are simulating an intrinsically chaotic system. Just as randomness in the real world makes the weather less and less determined the further one looks into the future, so very small changes in the forecasters' input data will vary forecast predictions. The input data that meteorologists use is the observed weather now (or nearly now), as relayed in from thousands of weather stations and satellite observations. These observations are only known to limited accuracy. For example, temperatures in reality may perhaps differ by ±0.1°C from the reported values, precipitation by ±0.1 mm from the reported values, humidity ±2% and so on. To calculate the effect of this, meteorologists make an 'ensemble' (or collection) of computer predictions, each computing run using slightly different input data varying within the limits of its uncertainty. They look at the resulting variation of predictions. If say 80% of the runs predict rain tomorrow afternoon but 20% say it will be dry then the forecast will say there is an 80% chance of rain. That doesn't mean it will rain for 80% of the afternoon. In fact the forecast statistic itself doesn't distinguish between continuous rain and passing showers. Clearly the forecast randomness increases with the range into the future of the forecast, which is why tomorrow's forecast is more certain than that for 5 days' time. Even tomorrow's forecast won't be completely certain because the atmosphere is intrinsically fickle.

The cause of randomness isn't always obvious. For example on a showery day, rain-radar will show isolated showers of perhaps a kilometre or two in extent drifting across the countryside and fading out. Is their location intrinsically random or might they be better predicted by entering more information into the forecast computer? The presence of showers can affect by thousands the number of people attending major sporting events. At present it seems that forecasters are getting better at predicting the location of such showers on a short time-scale, so some of the randomness is certainly an artefact of our ignorance. However it is costing millions to create the facilities to run fine-scale forecasting models on big enough computers. Even if future affordable computing improves, as it will, there will be a limit caused by the intrinsic randomness of the atmosphere. Meteorology will always remain a probabilistic science.

Correlations

All books on randomness start off by discussing the randomness of independent events. There are plenty of real world examples but not all forecasting randomness is like this.

One feature of independent events is that to find the probability of two events both occurring you can multiply their probabilities. Suppose the chances of being caught in a shower today and tomorrow are both given as 20%, or 1 in 5. The chances of being caught on both days is ${}^{1}/_{5}\times{}^{1}/_{5}$ which is ${}^{1}/_{25}$ or 4%. You can look at it the other way around. The chances of it staying dry when you are out on each day is 80% and hence the chances you will not get caught in a shower on both days is ${}^{4}/_{5}\times{}^{4}/_{5} = {}^{16}/_{25} = 64\%$. The probabilities of all possibilities must add up to 100% and in this case the 'missing' 32% represents the probability of being caught in a shower on 1 of the days only.



Typical Scottish rain events showing mesoscale distribution

The Met. Office and other bodies give forecasts in hourly slots for today and tomorrow. If their forecast shows a 50% chance it will rain between 1.30 and 2.30 and a 50% chance between 2.30 and 3.30 are these two chance estimates independent? If the weather predicted is isolated showers in the aftermath of a cold front, then perhaps 'yes'. In the UK, the majority of rain is from 'mesoscale' events, as the meteorologists say, activities that span typically a few tens of km. On the other hand, if the weather predicted is an extensive covering of stratus cloud that will take hours to pass then if it starts to rain in the first hour it's pretty likely to continue to rain into the second hour, so the predictions are not independent. Weather systems that bring continuous rain usually take more than an hour to pass. The forecast won't tell you this. It's another example of the fact that you can't just look at probability numbers in isolation but need to know something about what they refer to and where they come from. If two probabilities do not refer to independent events then we can't multiply them to get the probability that both events will happen.

I'll give one more example of the use of Met Office figures. Quite common in the summer in our location is the hourly forecast prediction that the probability of rain is 10% for any hour of the day. Personally I take it that if I have an outdoor activity planned for a few hours it's likely to be dry. Taking the figures at face value, what is the chance over a 10 hour period, say 9 am to 7 pm, there will be some rain sometime? The best way to answer this is to recognise that the probability of it being dry in any hour is 0.9. On the assumption that any rain is likely to be showers forming 'at random' and independent of each other, then the probability that all 10 hours will be dry is $0.9^{10} = 0.35$. So there is a 35% chance that it will be stay dry all day and a 65% chance that at least one hour will bring some rain. Of course, there are assumptions behind this figure and the 10% per hour is itself a rounded figure, not a precise one. In forecasts beyond two days, the Met Office provide 3-hourly forecasts. Note that a 10% chance of rain in a 3-hour slot is not the same as a 10% chance of rain in a 1-hour interval.

Don't slate the meteorologists for not being right all the time. Even our grandchildren aren't going to grow up with completely accurate forecasts, whatever the computing power available in the future. Weather has a random component – end of story.

Climate is a different issue

Predicting climate is not the same as predicting the weather. We can't tell in the summer or even in November if it will snow on Christmas day but we can tell that winter as a whole will be colder, wetter and windier than summer. Climate has predictable features. Other aspects of life are like that too. I don't know if I'll catch a cold in a fortnight's time but I do know that many more people will catch colds or the flu in the winter than in the summer. Some 26,000 people were killed on the roads in EU countries in 2013, more during the summer than the winter, and it is almost certain that some 24,000 will be killed in 2014, though the year is only half-way through as I write this*. I can't predict whether I will be one of them except to assign a probability. (On a population basis alone, 24,000 out of a population of 742 million makes it a chance of 1 in about 31,000). The point I'm making is that individual events may be unpredictable but bulk trends follow different rules.

The question everyone is asking is 'can changes in climate be predicted'? We don't know if this coming winter will be warmer than average or colder, or just average, but climate is defined as the 30 year average of weather. Climate is predictable. Climate is determined by slowly changing influences on a global scale such as land usage, ocean currents, cloud cover, atmospheric composition. Climate predictions come with the forecasters' randomness caused by the approximations in different climate models. All predictions won't be the same but the spread of predictions is very likely to include what will actually happen. All climate models on the international stage are serious undertakings based on the science of how the weather system works. You may be inclined to put more weight on some predictions than others but the fact is that climate change predictions are now accepted as telling us what is likely to happen based on a range of scenarios – accepted by governments, business, humanitarian agencies and other NGOs, the military, insurance companies and the great majority who have looked at the evidence. In short, randomness in the weather does not imply randomness in climate.

John S. Reid

* Updating the information above - the EU Commission statistics showed a trend of diminishing numbers of road accident fatalities since 2000 but in fact the fatalities in 2014 and 2015 each equalled those in 2013.