Math 459 Lecture 5

Todd Kuffner

This week

- Philosophical issues in statistics.
- Principled statistical motivations for Bayesian analysis.

Concepts in the Philosophy of Science

Epistemology how do we know what we know?

Induction reasoning that something is true in general based on observing several examples

Deduction reasoning that something is true based on logical implications of some assumed premises

Falsification evidence can only be used to disprove a hypotheses; supporting evidence does not verify that something is true

Empiricism knowledge is obtained by observing evidence in the natural world

Falsification and Science

Principle of falsification:

A scientific hypothesis must be falsifiable.

Example Hypothesis: All birds can fly.

A hypothesis which cannot be proven false is not scientific.

Example Hypothesis: The reason someone falls in love with another person is that invisible ghosts send undetectable signals to that person's brain telling them to love that person.

Probability Meets (Frequentist) Statistics

Statistical Hypothesis Testing Fisher, Neyman, Pearson (1920s, 1930s)

A hypothesis gives rise to a probability distribution that would result if the hypothesis is true.

- **Reject** if the observed data does not support they hypothesis
- **Fail to Reject** if the observed data supports (at least approximately) the hypothesis

'Support' is quantified using probability.

Timeline of Probability

Origins 1650s in France (pen pals Fermat and Pascal); gambling

Improvements Jacob Bernoulli (1713); Abraham de Moivre (1718): LLN and early CLT

Bayes 'An Essay towards Solving a Problem in the Doctrine of Chances' (1763)

Breakthroughs Gauss (1795; least squares); Laplace (1812; MGF-'Laplace transform', CLT)

Axioms Kolmogorov (1933): probability axioms

Axioms

Let Ω be non-empty set; call a set F of subsets of Ω a sigma-algebra (or sigma-field) on Ω if $\Omega \subseteq F$, F closed under complementation and countable unions.

Let P be a function $P : F \to \mathbb{R}$

1. $P(A) \geq 0 \ \forall A \in \mathcal{F}$, F
2. $P(\Omega) = 1$ (the probability of a certain event)
3. Countable additivity: $A_1, A_2, \ldots$ a countably infinite sequence of pairwise disjoint sets, each in F, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

A P satisfying (1-3) is a probability measure and (Ω, F, P) a probability space.

‘The Probability Calculus’

The Axioms together with some rules/definitions:

Conditional Probability definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

Comments

- other axioms/rules exist, but those above generally accepted by everyone (frequentists and Bayesians)
- given some probabilities as inputs, ‘probability calculus’ tells us how to compute further probabilities
- however, other than saying $P(\Omega) = 1$ and $P(\{\emptyset\}) = 0$, they give no guidance about *initial* assignment of probabilities
- for how to assign initial probabilities, need to discuss interpretations of probability

What does probability mean?

Three interpretations:

Classical/Logical measure objective evidential support; ‘in light of the relevant astronomical data, it is *probable* that Earth will experience a major electromagnetic pulse in the next 20 years’ (e.g. geomagnetic solar storm of 1859 ‘Carrington Event’, solar storm of 2012)

Bye-bye Electricity

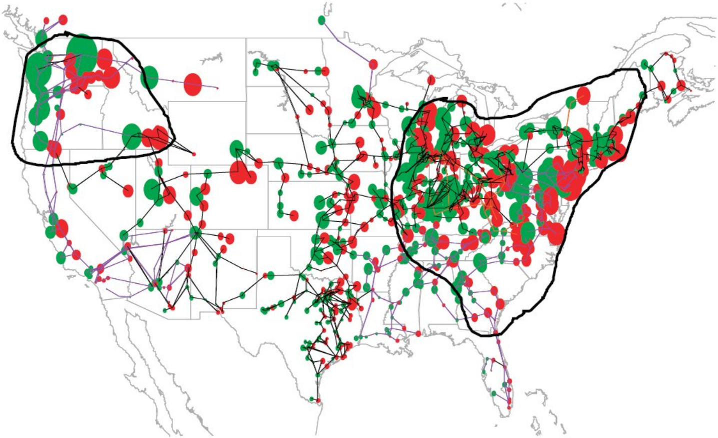

FIGURE 7.1 Scenario showing effects of a 4800 nT/min geomagnetic field disturbance at 50° geomagnetic latitude scenario. The regions outlined are susceptible to system collapse due to the effects of the GIC disturbance; the impacts would be of unprecedented scale and involve populations in excess of 130 million. SOURCE: J. Kappenman, Metatech Corp., "The Future: Solutions or Vulnerabilities?," presentation to the space weather workshop, May 23, 2008.

Classical Probability

Laplace, Bernoulli, Leibniz, etc.

- in the absence of any evidence, or if evidence is balanced among possible outcomes (*symmetrically balanced evidence*), then

all outcomes should be treated as equally likely

- this presumes finite outcome spaces
- need to be able to determine the entire set of possible outcomes
- e.g. irrational-valued probabilities would not be possible (e.g. $1/(3\sqrt{2})$)
- related to 'principle of indifference' (Keynes)

Comments

Extension to countably infinite spaces using *principle of maximum entropy* (argued by Jaynes):

- given a discrete assignment of probabilities $P = (p_1, p_2, \ldots)$, entropy of P defined to be

$$-\sum_i p_i \log p_i.$$

- entropy measures lack of ‘informativeness’ of a probability function
- more concentrated probability $\Rightarrow$ less entropy
- more diffuse probability $\Rightarrow$ greater entropy
- Principle of Maximum Entropy: select, from the family of all probability functions consistent with our background knowledge, the function that maximizes the entropy

Special Case: choosing uninformative prior over finite set of possible outcomes: choose ‘flat’ (uniform) probability

Further comments

Criticism if you are ignorant about the relative occurrence of different outcomes, the 'principle of indifference' is still extracting information from a state of ignorance

Response at least this interpretation of probability gives us a way for codifying ignorance into our epistemological process; i.e. it clarifies that *anything other than assignment of equal probabilities* indicates ignorance is not complete

Counter if ignorance is complete, better to not assign probabilities at all, or to assign imprecise probabilities (i.e. intervals of probabilities)

Second Interpretation

Subjective degree of belief; 'I am not sure if St. Louis will get another NFL team, but *probably* it will'

Often analyzed in terms of betting behavior, e.g. de Finetti:

> *Your degree of belief in E is p if and only if p units of utility is the price at which you would buy or sell a bet that pays 1 unit of utility if E occurs/is true, and pays 0 units if E does not occur.*

- presupposes that for any E, there is exactly one price–this is that person's *fair price* for the bet on E

Dutch book (against an agent)

A series of bets, each of which is acceptable to the agent, but which collectively *guarantee her loss*, however the world turns out.

- if your subjective probabilities conform to the probability calculus, then no Dutch book can be made against you
- then your probability assignment is said to be *coherent*

The Dutch book argument shows that beliefs about probabilities obey the probability calculus.

Dutch book example

Suppose a bookie sets this system of odds:

Book	**Odds**	**Probability**	**Bet**	**Payout**
E occurs	Even	$1/(1+1) = 0.5$	20	$20 + 20 = 40$
E doesn't occur	3 to 1	$1/(1+3) = 0.25$	10	$30 + 10 = 40$
		$0.5 + 0.25 = 0.75$	30	40

Probability implied by thesse odds sums to 0.75.

- if gambler bets 20 on E and E occurs, gambler takes 40 (original 20 plus 20 more)
- gambler also makes bet of 10 on not E; if 'not E' occurs, gambler also takes 40 (original 10 plus 30 more)
- gambler played $20 + 10 = 30$ to play, but no matter what happens, she wins 40

Interpretation

No matter what happens, the gambler making this bet has a profit of 10.

- the gambler can make a Dutch book against the bookie
- this is because the bookie has prior beliefs about the odds which do not obey the probability calculus! $(P(\Omega) = 0.75 \neq 1)$

Bayesian probability usually has some requirement that beliefs are rational and coherent.

- 'rational' is not a consensus definition

Third Interpretation

Frequentist objective concept that applies to various phenomena/systems in the world, independently of what anyone thinks; 'a particular tree will *probably* die within the next 200 years'

Finite Frequentism:

> *the probability of an event A in a finite reference set B is the relative frequency of actual occurrences of A within B*

Classical count all possible outcomes of a given experiment

Frequentism count actual outcomes

Generalization to Infinite Reference Sets

Problem: relative frequences must be *relativised* to some reference set.

Solution: two axioms concerning hypothetical frequentism

Convergence the (hypothetical) limiting relative frequency of any event exists (long-run relative frequency)

Randomness the limiting relative frequency of each event in a (hypothetically infinite) collection/sequence of events is the same in any infinite subsequence of that collection

Collections are abstract objects with no empirical substantiation, but frequentists (e.g. von Mises) say they can explain the stabilities of relative frequencies in actual sequences of outcomes of a repeatable random experiment.

Karl Popper (1902-1994)

A probability p of some outcome is a propensity of a repeatable experiment to produce outcomes of that type with limiting relative frequency p.

- e.g. we say a coin has probability 1/2 of landing on heads when tossed; this means we have a repeatable experimental set-up (tossing a coin)
- this experiment has a propensity to produce a sequence of outcomes in which the the limiting relative frequency of heads is 1/2

Bayesian Statistics

- Probabilities are interpreted as 'relative beliefs' (subjective probability)
- <u>However</u>, Bayesian statistics much broader than just an interpretation of probability

Back to the philosophy of science: how do we use evidence to do hypothesis testing?

What is a p-value?

The probability that, *if the null hypothesis is indeed true*, we would have observed data which contradicted the null hypothesis by *at least* as much as the data we actually observed.

- probability here is long-run frequency under hypothetical repeated sampling (frequentist probability)

How do we find p-value?

- say $H_0 : \theta = \theta_0$, $H_1 : \theta > \theta_0$;
- use test statistic $T(X^{(n)})$; observe $X^{(n)} = x^{(n)}$, compute test statistic, e.g. $t(x^{(n)})$

$$p = Pr_{\theta_0}(T(X^{(n)}) \geq t(x^{(n)}))$$

- viewed as function of θ, $p(\theta)$ called a **significance function**

Controversy 1

John P.A. Ioannidis (2005), ‘Why most published research findings are false.’

Replicability
: replicable findings are those which are found on multiple samples from the population of interest

PPV
: positive predictive value: post-study probability that a research finding is true, after it has been claimed to be true on the basis of statistical significance

Explanations:

1. Power (small sample sizes)
2. Bias (usually selective reporting of findings that are publishable)
3. Bad statistics

Controversy 2

In 2014, the journal *Basic and Applied Social Pscyhology* banned the ‘null hypothesis significance testing procedure’.

> *(The p-value) fails to provide the probability of the null hypothesis, which is needed to provide a strong case for rejecting it*

- March 2015 ISBA (International Society for Bayesian Analysis) bulletin responds

Bayesian View of Hypothesis Testing

Hypotheses are like parameters, we model them as random variables.

- attach a prior probability to each hypothesis
- observe data
- update beliefs to posterior probability of each hypothesis

Advantage: this is how people naturally want to think about combinging testing and probability

Hypothesis Testing in Statistics

Does rejecting the null mean it is ‘disproved’?

Does rejecting the null mean that the alternative is true?

Does failure to reject the null mean that the null is true?

What are statistical hypothesis tests really?

If they are not considered proof or disproof, what are they?

- statistical inference provides a framework for making statistical decisions using observed data
- making a statistical decision about a hypothesis does not mean it has been proven or disproven

Next time: What do we mean by decisions?